"Modern peto" and "revised peto"

- My " Revised peto"

  1. Conceptually similar to ARPC Peto, but uses only meso claims to estimate exposed population as of end of calibration period

  2. Forecasts assume increasing lifespans, a la Social Security

  3. Uses very detailed, 3-dimensional "cells"; then uses a GAM smooth

- My "Modern peto"

  1. Fits a mortality model to national data, similar to UK Health Safety Executive

  2. Fitted model plus Social Security lifespan forecasts yields forecasts of national U.S. deaths, by year and birth cohort

  3. Find "propensity to sue" by birth cohort in calibration period, then prorate forecasts

Revised Peto

- Agree that controlling for available exposure info does not make a huge difference
- Basic demographics (i.e. age profile of exposed population) is important
- Basically, most workers and spouses are exposed around age 18
- Still worth checking exposure info
- Steps in my process

  1. Deal with logically impossible or suspect exposure data
     E.g., year of first exposure precedes birth

  2. Multiple imputation of missing year of first exposure, duration
     f(birth year, prob(paid), secondary, child, foreign)

  Above steps are done for *mesos only*

Revised Peto 2

- Steps in my process (continued)

3. Build a contingency table
   60 age at first exposure
   82 birth years
   11 exposure categories

   60 x 82 x 11 = 54,120 cells

4. Eliminate structural zeros (e.g., year of first exposure after 1982)

5. Fill remaining cells with counts -- either "sampling zeros" or the number of meso claims from the calibration period

6. Fit a poisson model to the table; assume expected value in each cell is a smooth function of age at first exposure, birth year, duration, and their interactions

Revised Peto 3

- Steps in my process (continued)

7. Find the implied number exposed in each cell by inverting the Peto formula from the 1986 EPA report

   Basically same calculation done in ARPC Peto
   Only change -- I cap the risk at age 85+

8. Kill off the exposed population year by year using the Social Security estimates of future death rates

9. Calculate number of mesos each year using the modified Peto formula

Revised Peto -- Smoothing the contingency table

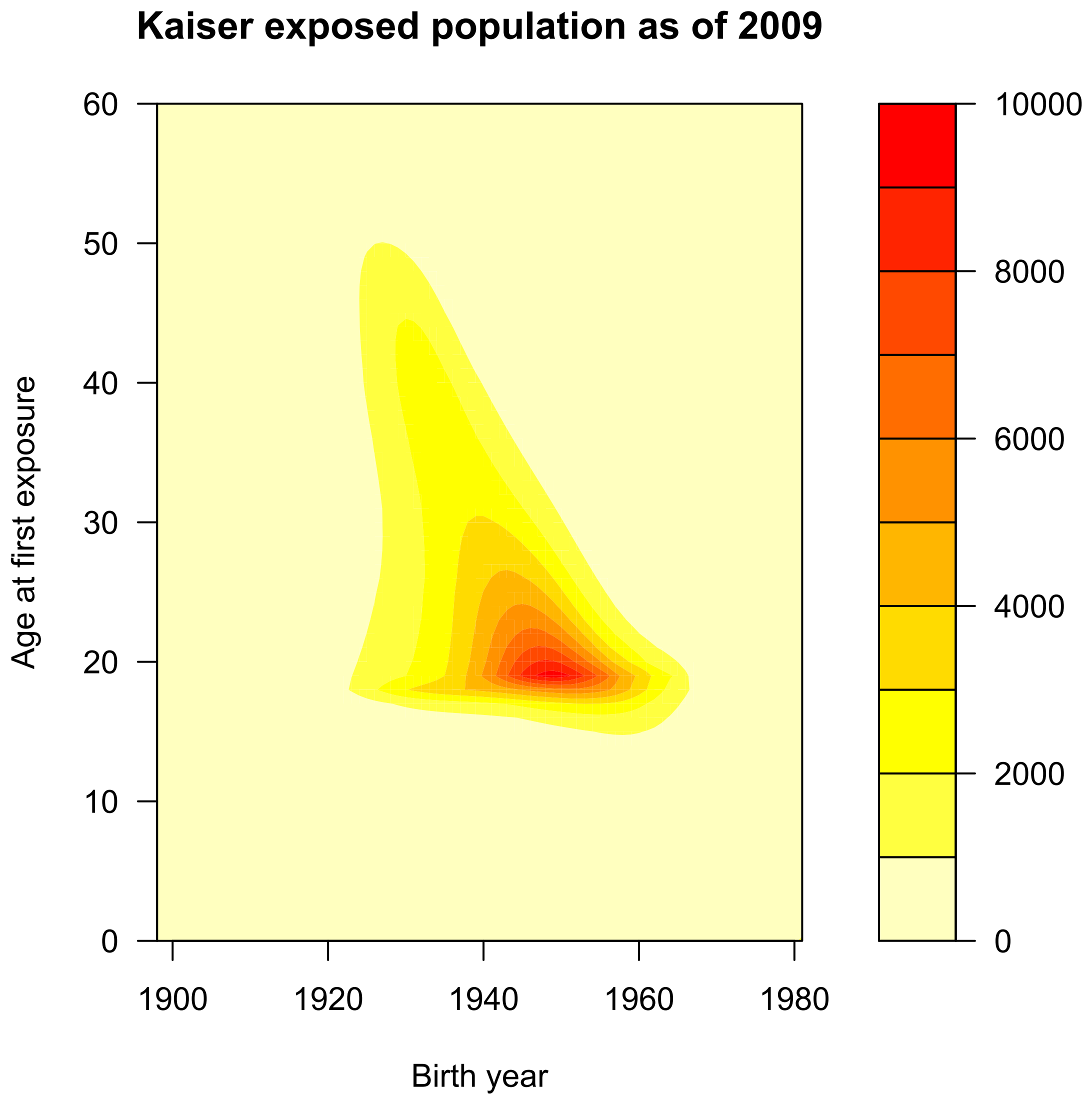
- All work done in R

- Generalized additive model -- poisson regression

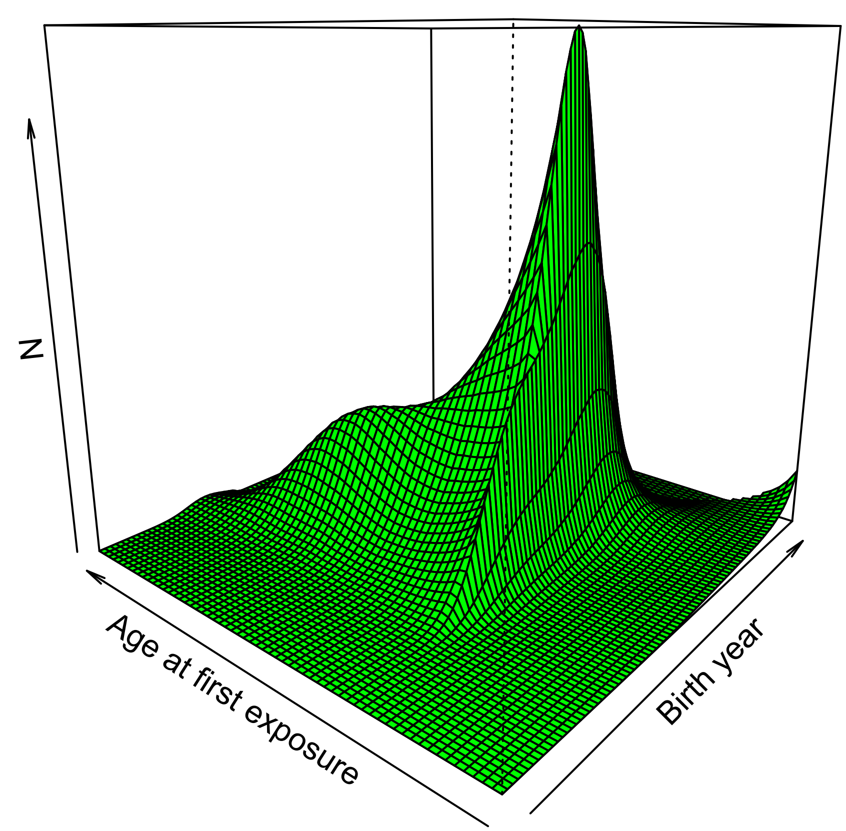
```
fit.old   = gam( y ~ te( birth.yr,exp.age,exp.dur.mean,by=arising),
                 data=m1[old,], family=poisson)

fit.young = gam( y ~ te( birth.yr,exp.age,exp.dur.mean,by=arising),
                 data=m1[young,], family=poisson)
```

Simon Wood -- mgcv package
Uses splines -- interaction accommodated with tensor products
Controls for over-fitting using generalized cross-validation

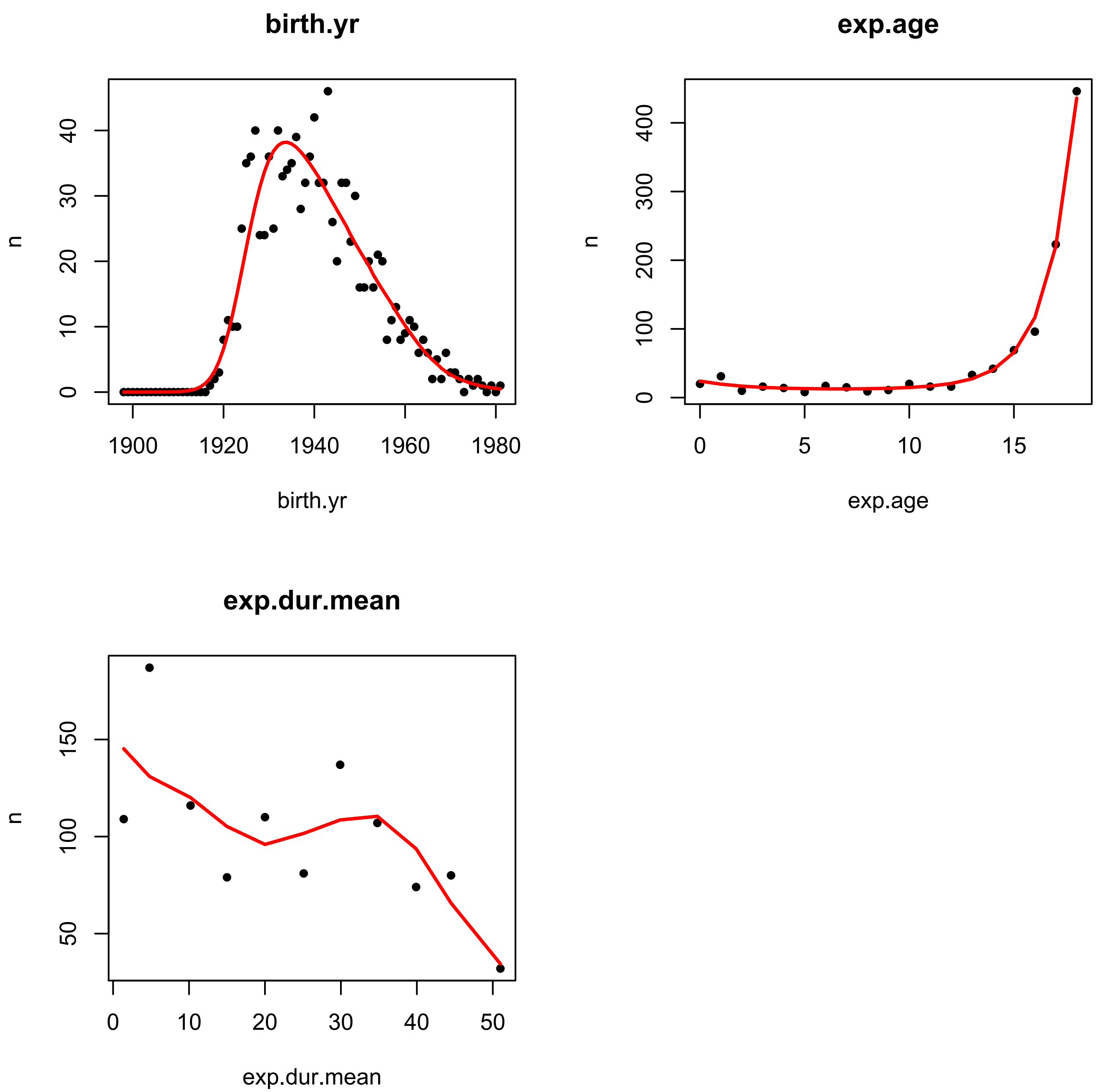
Note trick of smoothing separately for <= 18, > 19 -- allows fitting of sharp "ridge" of first exposure at age 18

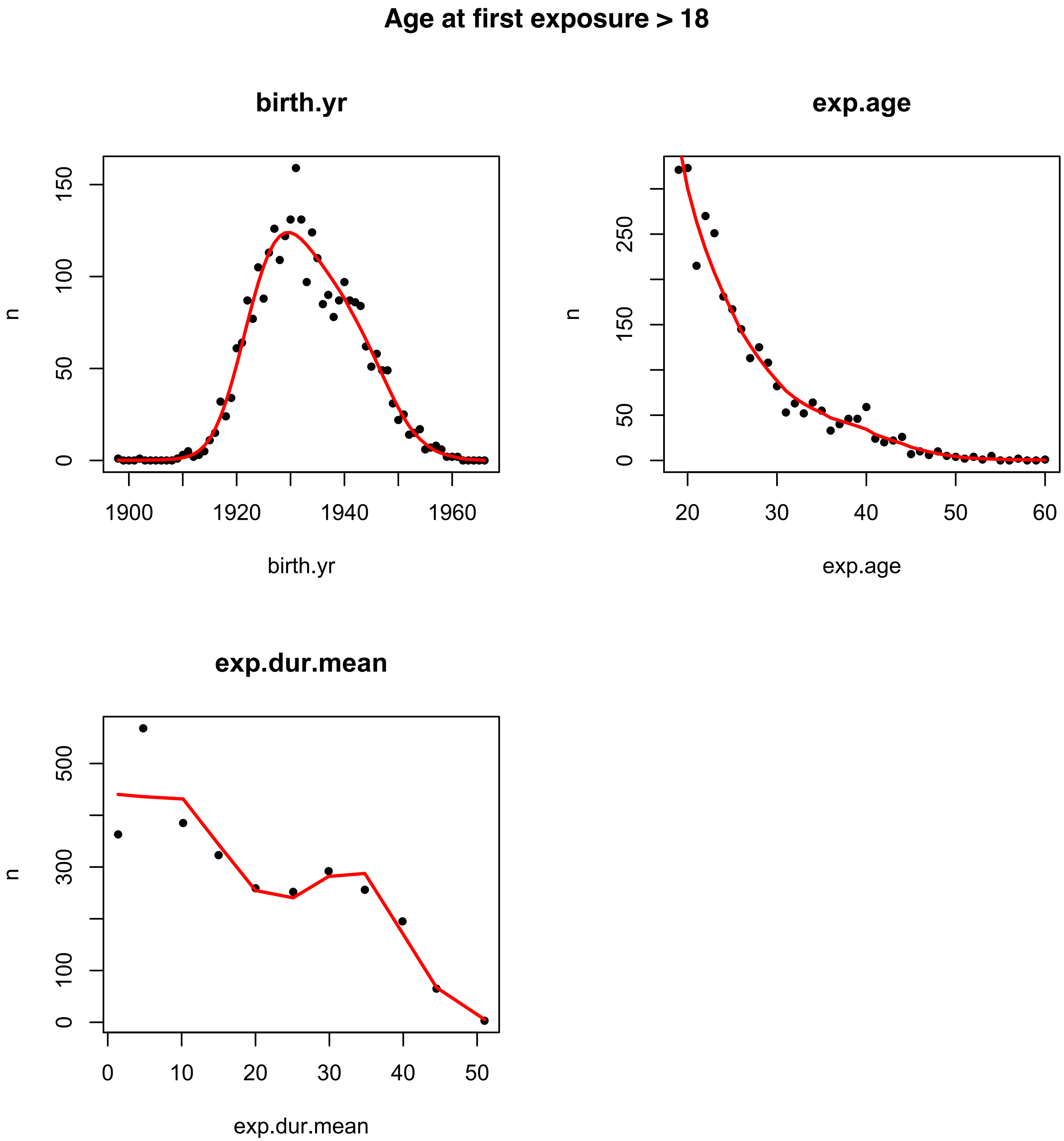

Kaiser exposed population as of 2009
Age at first exposure
Birth year
0
10
20
30
40
50
60
1900
1920
1940
1960
1980
0
2000
4000
6000
8000
10000

Kaiser exposed population as of 2009



Age at first exposure <= 18




Age at first exposure > 18
birth.yr
n
0
50
100
150
1900
1920
1940
1960
birth.yr
exp.age
n
0
50
150
250
20
30
40
50
60
exp.age
exp.dur.mean
n
0
100
300
500
0
10
20
30
40
50
exp.dur.mean

Revised Peto -- Additional comments and issues

- Multiple imputations done in R using "Mice" chained equations package
- I fill the contingency table with *valid* meso claims, not total meso claims

  One option is to just count meso claims by cell

  But instead ...
  I build a master dataset that has a field for the probablity of payment for each claim
  I sum these probabilities by cell

  Allows me to adjust for variation in payment probabilities by age at first exposure, birth year, etc.

- Other issues

  Dealing with changes in the exposed population within the calibration period
  Forecasting lung cancer, etc.
  Does one need to smooth at all?

Modern Peto

- Uses national data, creates a modern analog to "Nicholson curve"

- Assumes we have:
  - Total population, by year, sex, age
  - Total mesothelioma deaths (or diagnoses) by year, sex, age

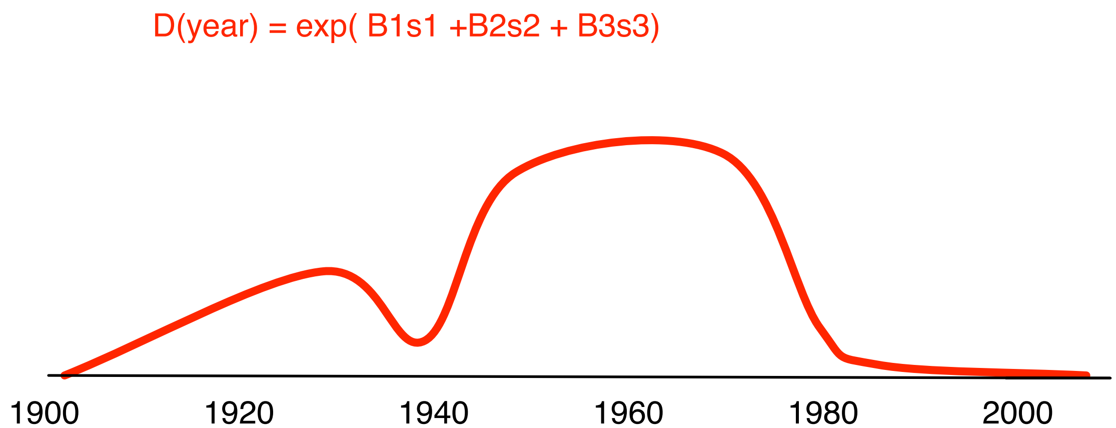
- Cohort-age models have not worked well

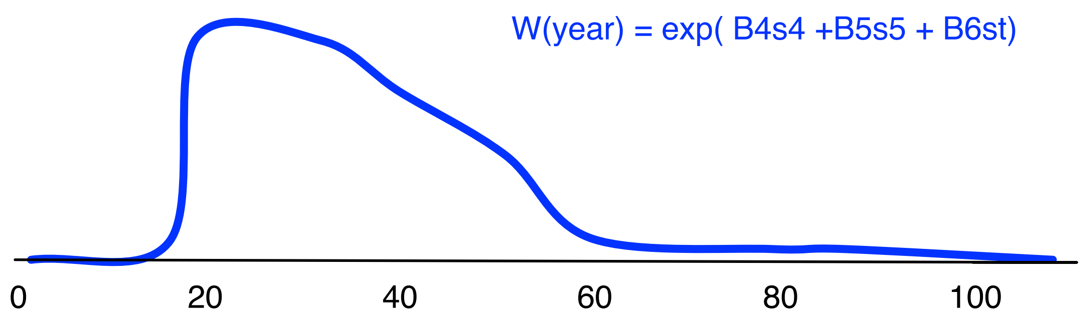
- Modern peto is designed to be more realistic

  W(age) is an index of exposure intensity as a function of age
    Roughly goes way up around age 18, flat from 18-30, then gradually declines
  D(year) is an index of population exposure as a function of year
    Roughly goes up in the 1930s, stays high through 1970, then declines

  Exact shapes of each index are fit empirically, to maximize a likelihood

Hypothetical indices, defined using cubic spline basis of log(index)





What is expected number of mesotheliomas in 2010 for cohort born in 1940, solely from exposure in the year 1960?

Number of deaths is poisson distributed, with mean alpha x D(1960) x W(20) x (2010 - 1980) ^ K

Peto model term -- latency of 20 years, then risk increases as a power of K

To get expected deaths in 2010 for this cohort due total exposure from 1940 to 2010, must cumulate the contributions from each year lived so far:

...
alpha x D(1960) x W(20) x (2010 - 1980) ^ K +
alpha x D(1961) x W(21) x (2010 - 1981) ^ K +
...

Fitting the modern Peto model

- Complex likelihood function, due to all the cumulating for different cohorts;
  highly nonlinear

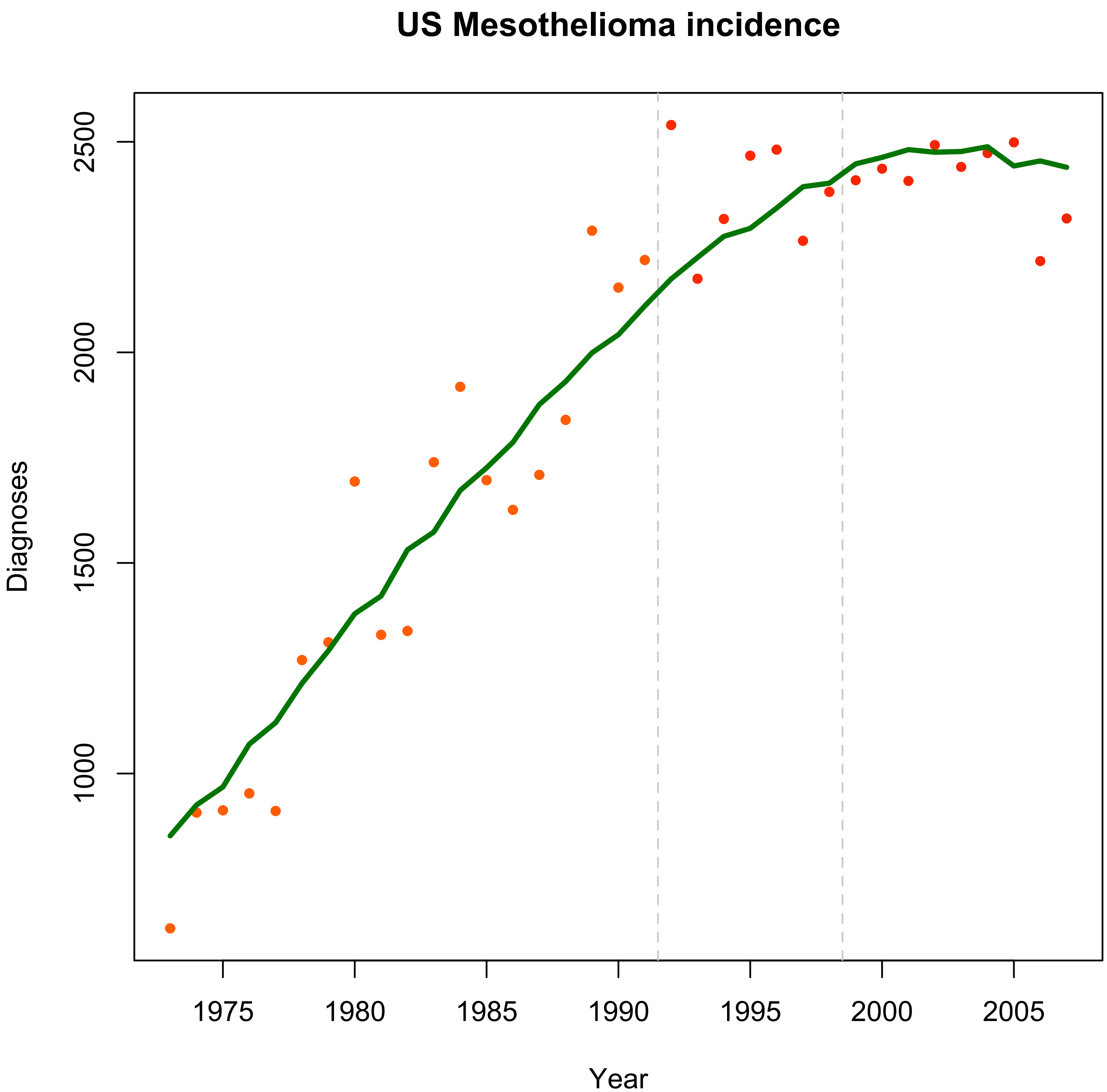
- Must use general purpose optimizers to maximize likelihood as a function of
  alpha, B1:B6, K (in the example)

  or use MCMC method

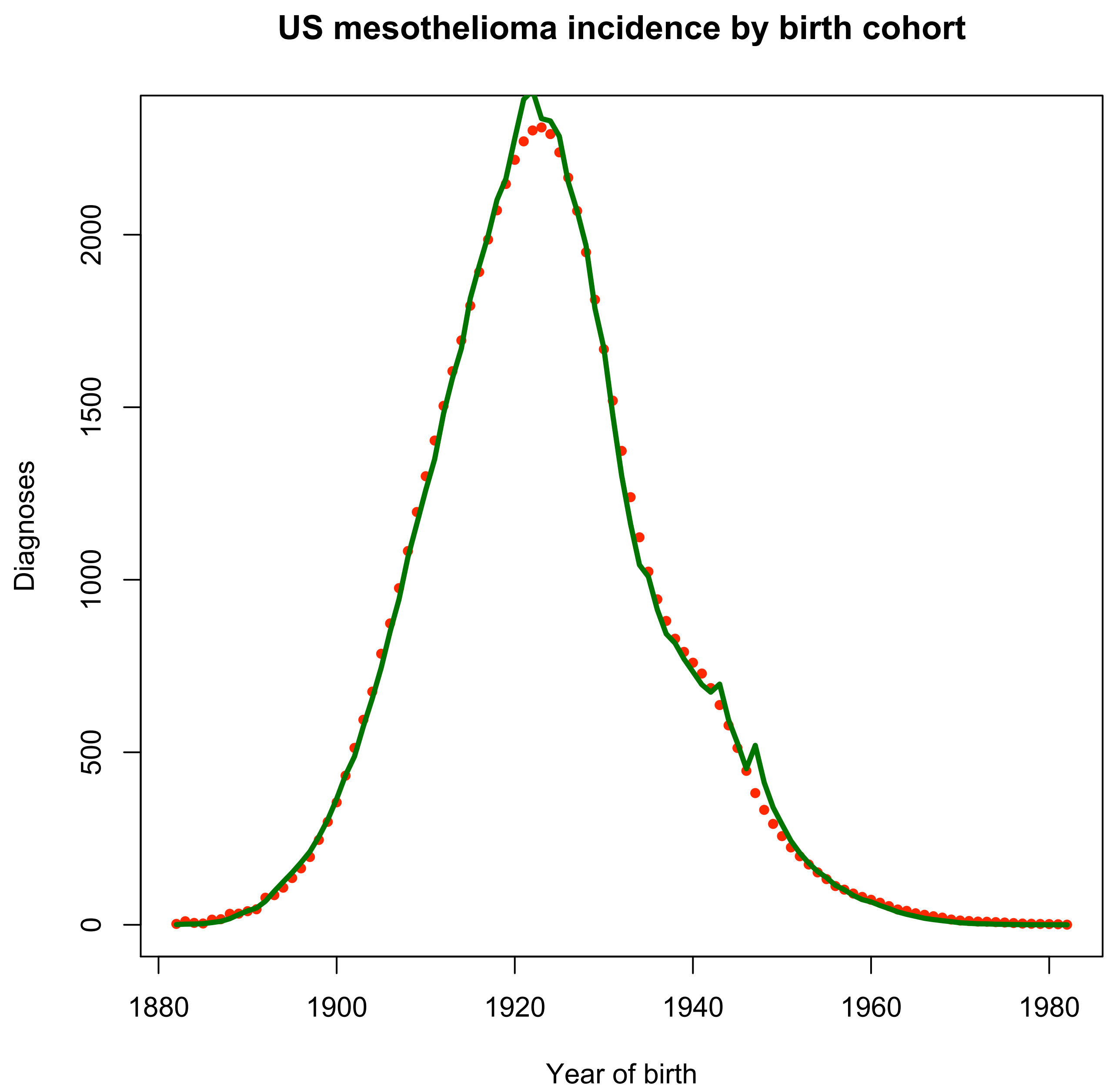
Hodgson, UK Health Safety executive, 2009	ad hoc optimizer
Tan, UK Health Safety executive, 2010	fminsearch() from matlab
	MCMC method
Gravelson, UK Actuary Society, 2010	unknown optimizer
Australian Actuary Socity, 2010 (?)	optim(), from R
Me	optim(), from R

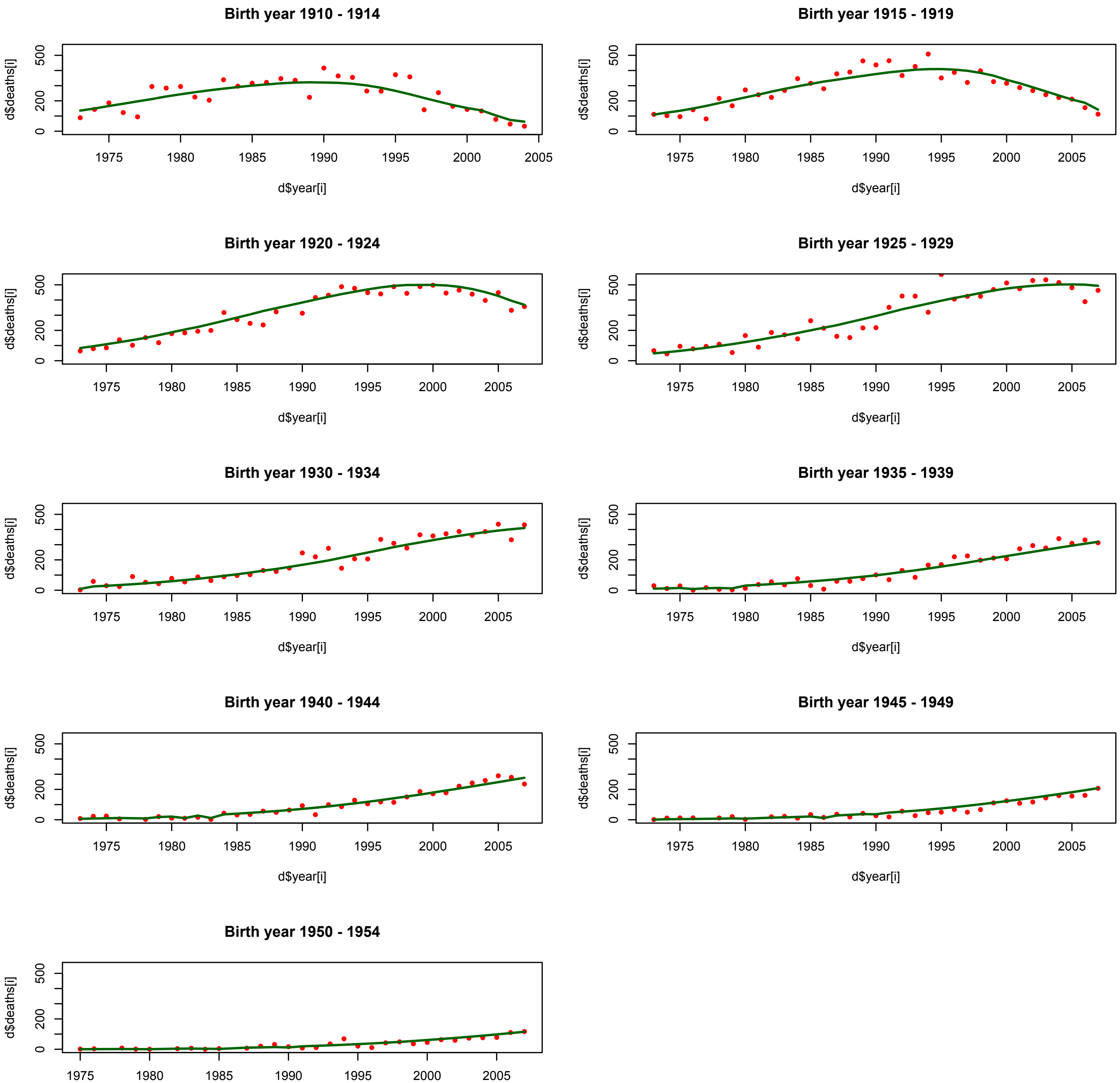
  I have also done some preliminary work with Rjags, but still exploratory

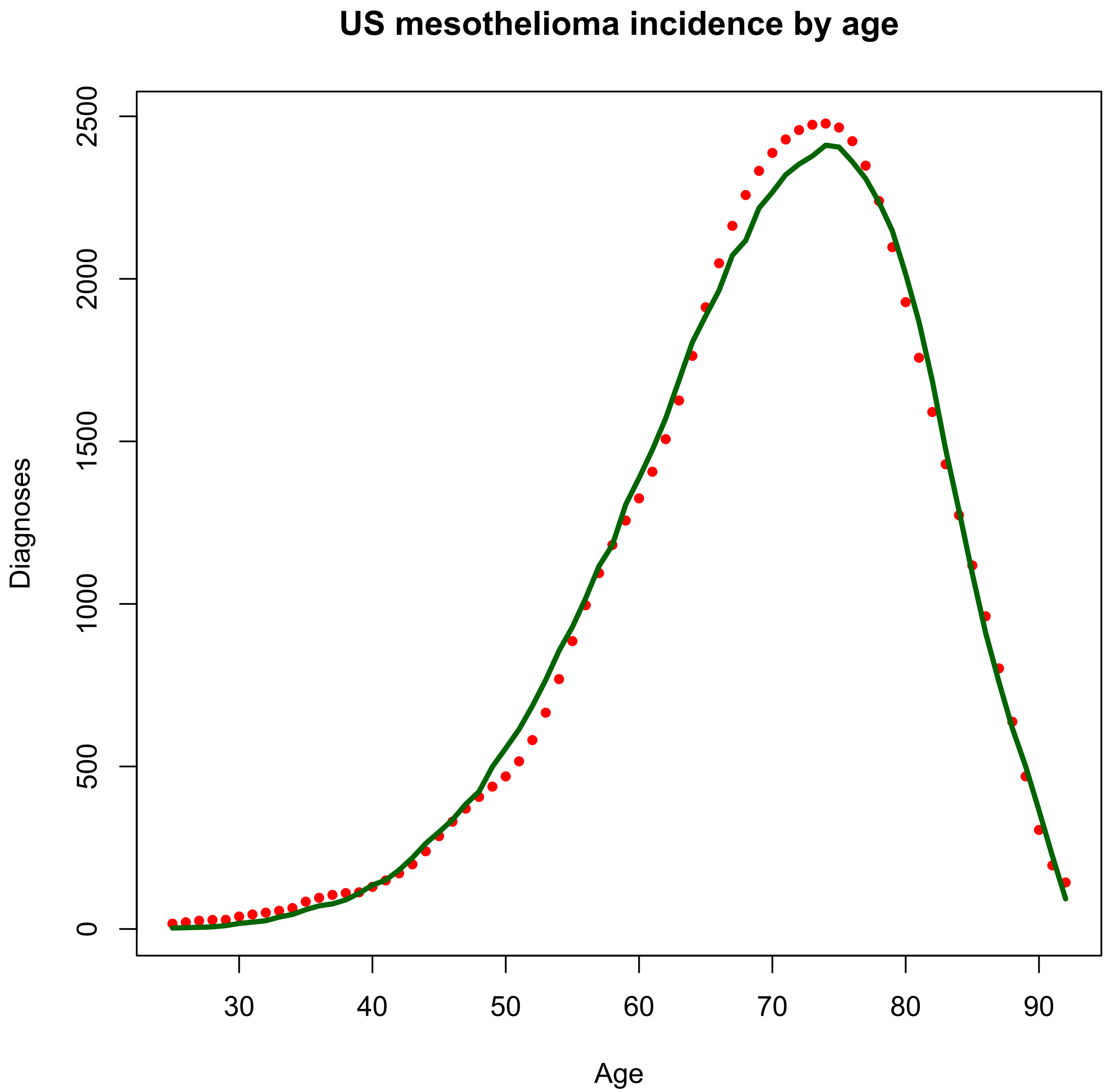
-



US Mesothelioma incidence
Diagnoses
1000
1500
2000
2500
Year
1975
1980
1985
1990
1995
2000
2005


US mesothelioma incidence by birth cohort
Diagnoses
0
500
1000
1500
2000
1880
1900
1920
1940
1960
1980
Year of birth


Birth year 1910 - 1914
Birth year 1915 - 1919
Birth year 1920 - 1924
Birth year 1925 - 1929
Birth year 1930 - 1934
Birth year 1935 - 1939
Birth year 1940 - 1944
Birth year 1945 - 1949
Birth year 1950 - 1954
d$deaths[i]
d$year[i]
0
200
500
1975
1980
1985
1990
1995
2000
2005


US mesothelioma incidence by age
Diagnoses
0
500
1000
1500
2000
2500
30
40
50
60
70
80
90
Age

Variations on the modern Peto model

- Form of the indices (exposure level by year, exposure intensity by age)

  I believe the Australians are using splines, as in my example

  I intend to move to splines, but to date have used variants of the idiosyncratic methods from Hodgson 2009

  Everyone else uses Hodgson-type methods, but with differing degrees of freedom

- Additional model terms used by various analysts

  Background death rate

  "Clearance half-life" -- related to some chrysotile asbestos studies, terms are insignificant in model fits

  Under-diagnosis of mesothelioma from 1970, but improving since then

  Flattening of risk at various ages, usually around 85-90

  Different constraints on how fast exposure declines to near-zero after 1970

- There are also different age ranges addressed -- 20-85, 20-90, etc.

Data for fitting the modern Peto model

- In U.S., we have all these different registers with wildly varying coverage --
  SEER 9, SEER 13, SEER 17, USCS

- I have not incorporated the latest USCS data point. May be more SEER as well.

- I scale USCS up to full population (already at like 98%), then use the most complete SEER data in any year, with a simple bias adjustment calculated from the years of overlap with USCS. I.e., USCS is the gold standard.

- Complete UK data is available online. (In fact, you can get many countries from WHO public data.)

- New data points have maximum impact at this point in time.

  People who were first exposed in 1980 are now typically in their 50s.
  They are the canary in the coal mine -- we would previously have started seeing measurable mesos for this group about now. How much are the counts declining?

General purpose optimizers

- There are speed, convergence, and "programming time" issues.

- I use the optim() methods that allow for constraints on each parameter -- helps avoid local maxima.

- General optimizers do not report "t-statistics" -- important to check marginal likelihoods from a fit -- may be flat, but "estimate" is at one end of constraining range

- Reasonable starting values and scaling help speed convergence

- Other speed-ups (not yet implemented by me) --

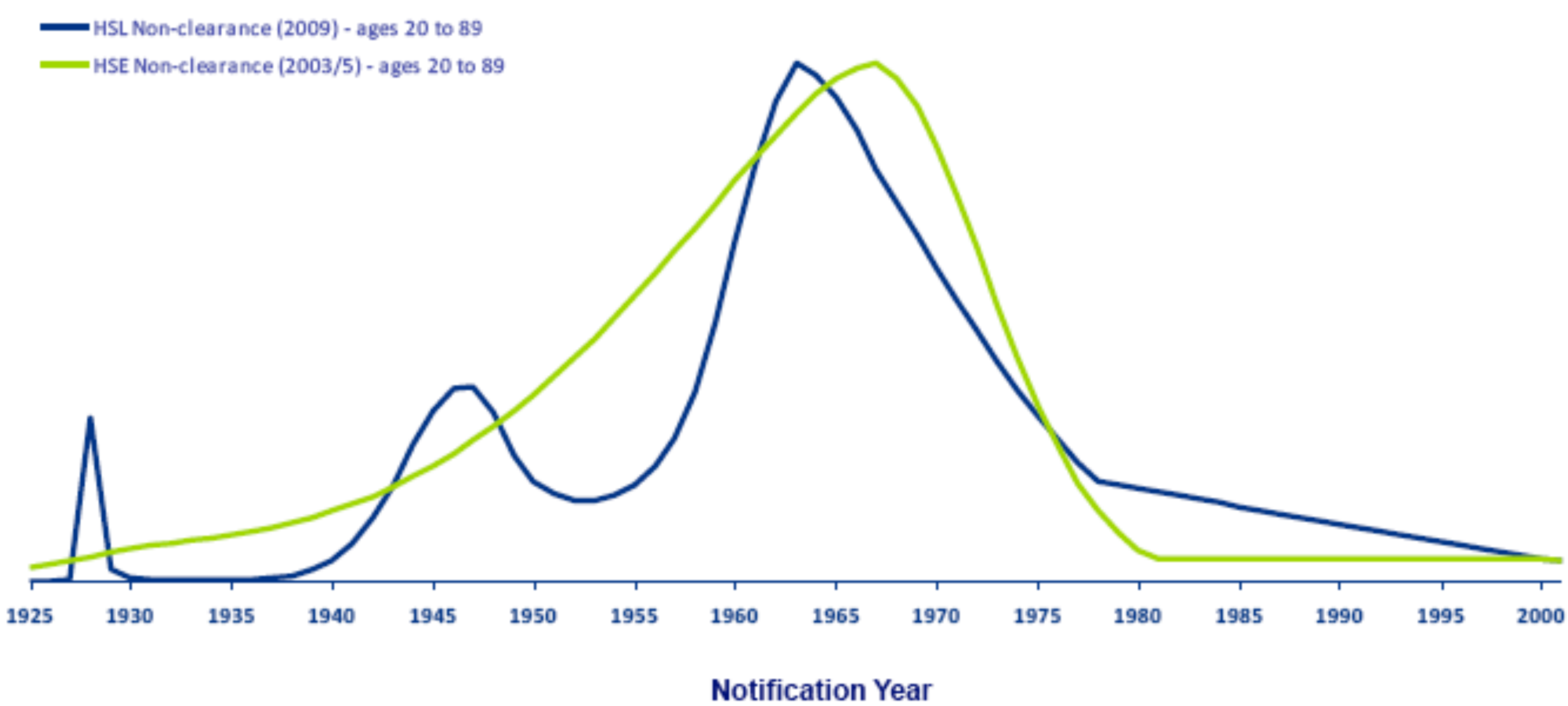
  Replace summing the likelihood contributions by looping with a matrix multiplication

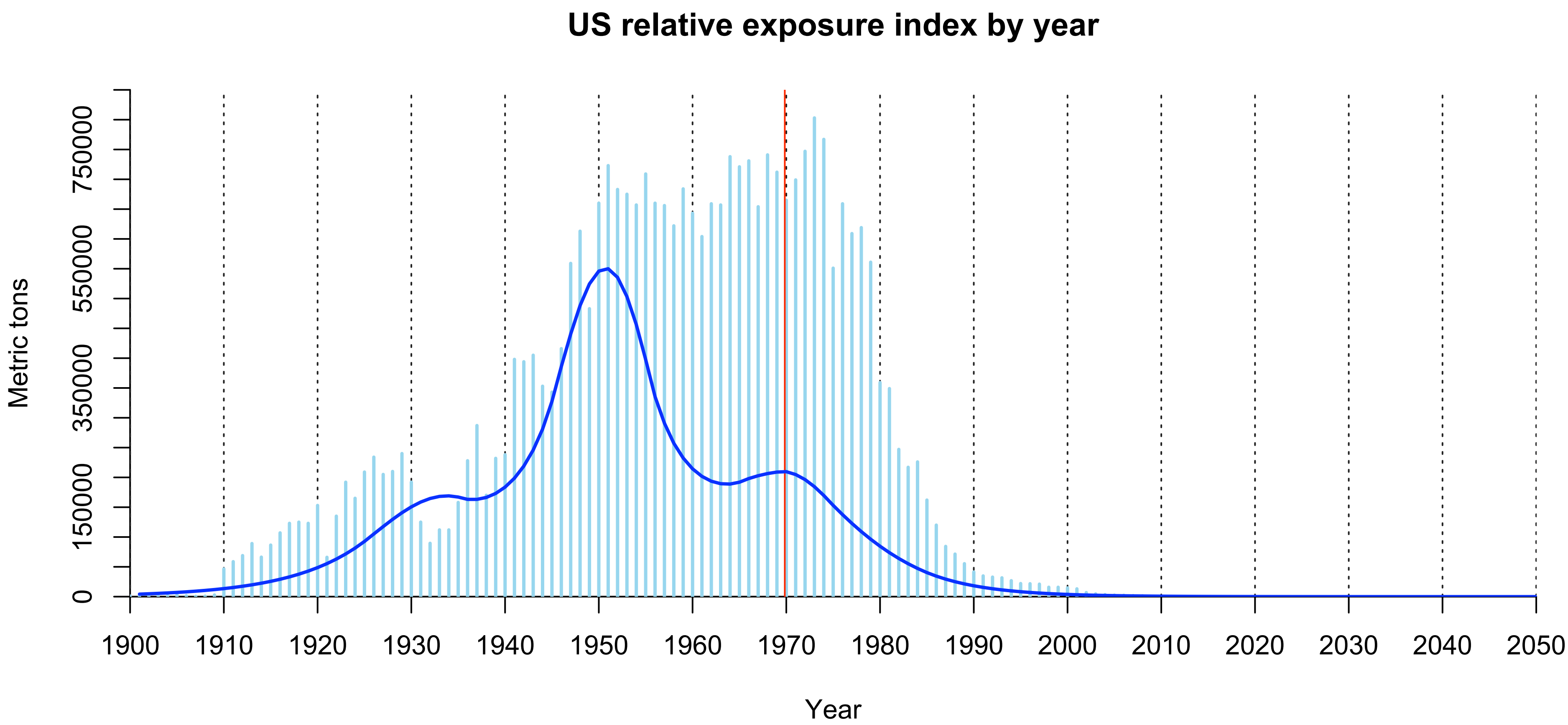
  Defining indices with splines should eliminate some costly table lookups

Where I'm at with modern Peto

- Need to improve housekeeping code to make it easier to quickly respecify model for different runs
- Plan to fit with updated U.S. data; also do a comparison fit with just SEER 9
- Plan to change index specification to splines
- Add speedups to optim() mentioned in previous slide
- Add model term for background rate
- Possibly expand age range (currently 20-90)
- Possibly do a model for women

HSE/HSL Exposure Assumptions





US relative exposure index by year
Metric tons
0
150000
350000
550000
750000
Year
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
2040
2050