# Staph Modelling write-up.

## Abstract

Attempting to fit a markov chain model to the staph carriage data. As yet, little success.

## Background

At 24 weeks,

## Model

Nasal swabs were taken from 1123 adults attending five Oxfordshire general practices. 571 were re-swabbed after one month then every two months for median two years. All *S. aureus* isolates were *spa*-typed. Risk factors were collected from interviews, general practice, and hospital records.

### Study population

Eligible participants were consecutive adults aged ≥16 years attending one of five Oxfordshire general practices (each a group of family doctors) in **the** Thames Valley Primary Care Research Partnership (all in the catchment area for the Oxford University Hospitals (OUH) NHS Trust), providing written informed consent*.* 200 participants were recruited from each general practice sequentially over December 2008-December 2009, in age/sex strata approximately representing the UK population. Recruitment was completed in each practice before starting in the next. To increase numbers of younger participants, students registering at one practice were recruited during the University Freshers’ week. For the first four general practices, we invited only those participants whose recruitment swab grew *S. aureus* to continue in the study. For the last practice and the students, we invited all participants to continue in the study.

Nasal swobs taken as described in (Ruth’s paper). Spa-typing was performed on all samples growing S. aureus and these *spa*-types were grouped into *spa-*Clonal Complexes (CCs) using BURP clustering, with a cost-threshold of 2. (in line with best practise in Ruth’s paper)

Outcome definitions

1. Loss of carriage

Confirmed loss of carriage was defined as the first of two negative swabs (or the first of two swabs without the previous *spa*-type for *spa*-level analyses). Single isolated negatives were ignored (given potential limited efficacy of self-swabbing). Participants with only their last swab negative were censored at the preceding positive swab. Thus loss analyses included only participants returning ≥2 swabs after the first positive to enable any loss to be confirmed. Loss rates over time were estimated using flexible parametric hazard models (26).

1. Acquisition

*S. aureus* acquisition was defined as positive growth (or a new *spa*-type) after confirmed prior absence. Thus if the first post-recruitment swab in recruitment-negatives grew *S. aureus* (or a new *spa*-type in recruitment-positives), this was not counted as acquisition but was presumed to represent a false-negative result at recruitment. Acquisition analyses therefore also included only participants returning ≥2 post-recruitment swabs. Since nasal evolution can produce small changes in repeat numbers, new *spa*-type acquisition was defined as having >2 differences from first positive swab (see Supplementary Table 1 for grouping).

All individuals were to be followed for two years (14 swabs in total) under the original protocol. If an individual did not return three consecutive swabs, no further swabs were sent. Following a protocol amendment, at two years further consent was sought for longer follow-up in those persistently negative or persistently positive (allowing single intermitted negatives) for *S. aureus* through to four years to enable longer-term rates of gain and loss to be estimated in those remaining at risk. 233 were followed up to 48 weeks.

To investigate *S. aureus* loss and (re-)acquisition, the 360 recruitment-positive individuals plus a further 211 without *S. aureus* at recruitment (82 from the last general practice, 129 students) were followed for a median (IQR) 2·0 (1·8-2·2) years, returning a median (IQR) 14 (11-15) swabs (range 1-20). Three (0.5%) individuals died and 121 (21%) participants were lost (25 (4%) did not return any swabs post-baseline, 53 (9%) missed returning three consecutive swabs and were removed from follow-up and 43 (8%) moved from the area or withdrew from the study) (Figure 1, Supplementary Figure 1). My dataset has 348 recruitment-postive, 198 recruitment negative (of whom on 59 are marked as students?). Followed for a median (IQR) 24 months (24-42). My set clearly diff, as everyone has at least 2 swabs. How can I work out the rest of these?

*S. aureus* grew from 4110 of 8190 post-recruitment swabs returned (50%) and was subsequently recovered from 92 (46%) recruitment-negatives, ten 18??? How have I got a different number here? (9%) at the first post-recruitment swab. Have I got wrong definition of first post recruitment? Should I be looking at timepoint =1 or 2?

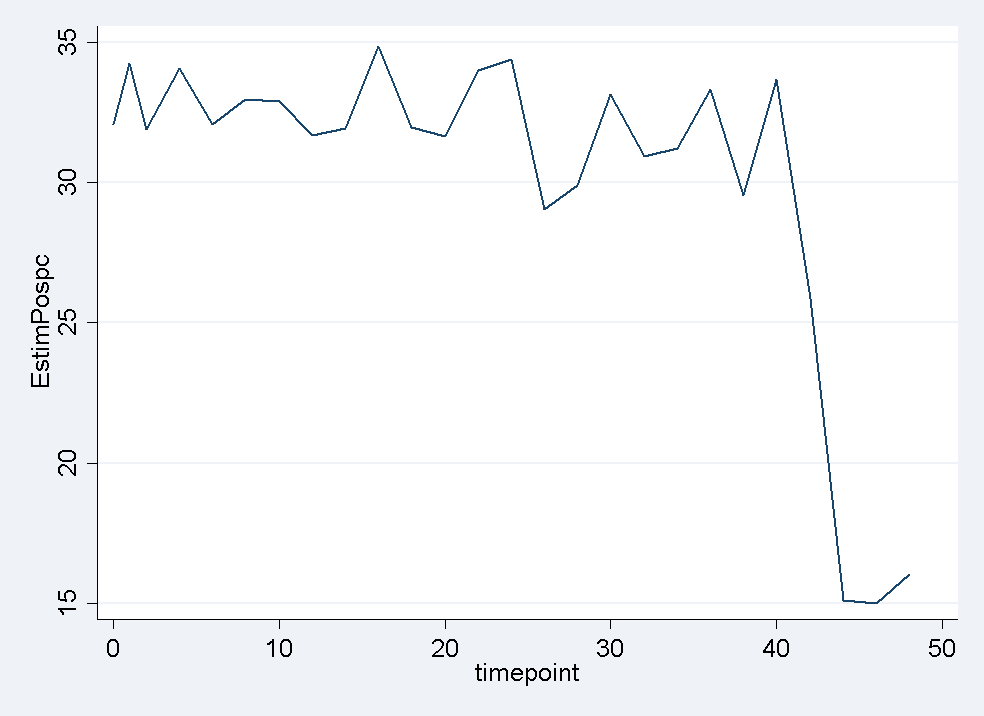
All *S. aureus* were *spa*-typed; of the 297 *spa*-types observed, 197 (66%) were only seen in one individual. The 297 *spa*-types formed 157 groups with ≤2 differences, 82 were singletons and 22 were excluded from grouping because they were too short (Supplementary Table 2).

Most spatypes seen in single individual over course of study = 7. 322 spa-types observed.

## Results

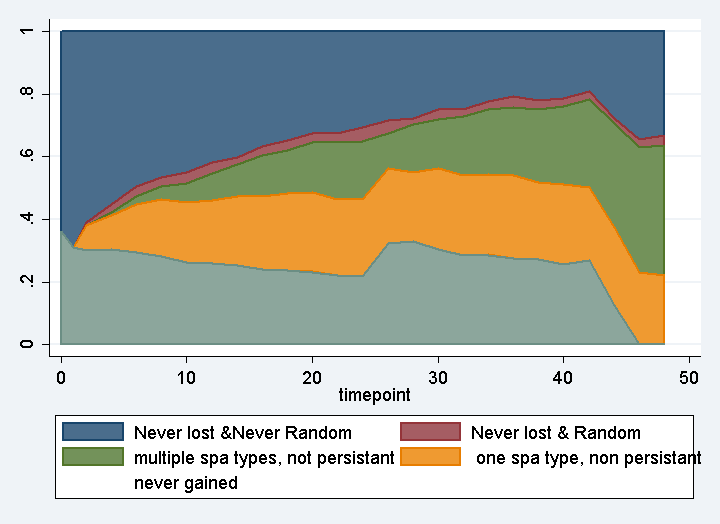
#### Equilibrium of % of people carrying staph

Results of pos/neg waited by %of initial dataset still following according to pos/neg intake state. Goes loopy post 24 weeks, as we filtered the subset and I haven’t worked out how we should account for that yet.

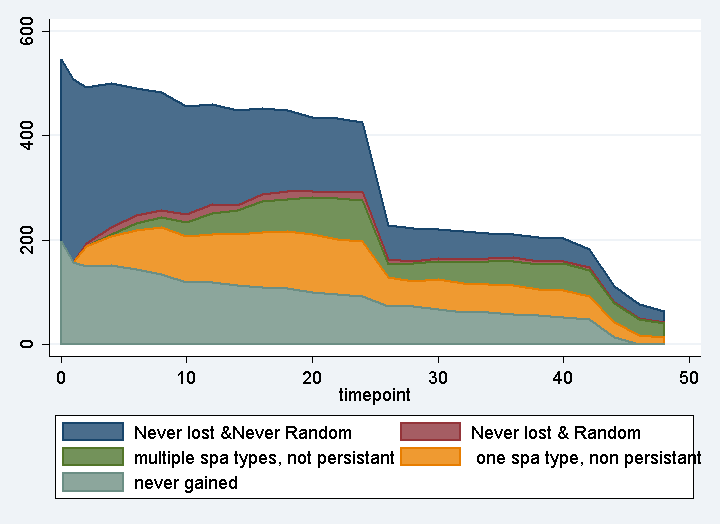


#### Looking at equilibrium patterns in moving between five types

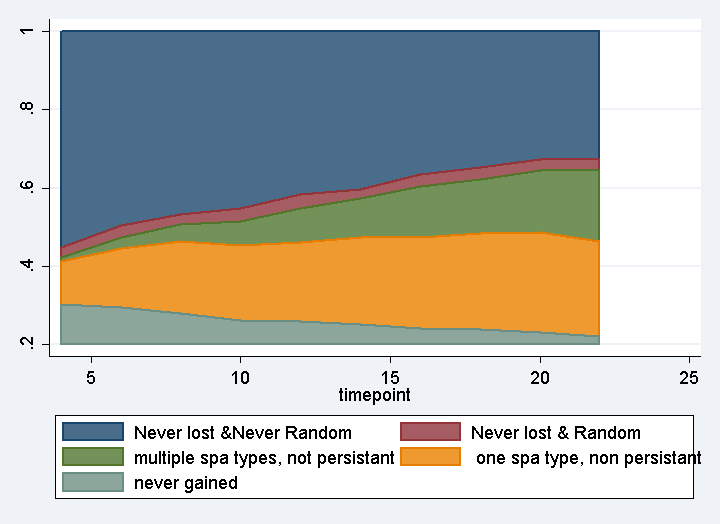
1. Never gained – no swab ever positive
2. One spa type, non persistant – at most one spa type detected (ignoring shifts of <=2 SNPs), but with confirmed loss (2 swabs negative in row) during study
3. Multiple spa types, non-persistant. Confirmed loss of all initial spatypes (2 neg swabs in a row) and presence of multiple spatypes > 2 SNPs.
4. Never lost and random. Persistant carriage of at least one initial spatype, and gain or loss of an additional spa type > 2 SNPs different.
5. Never lost and never random. Persistant carriage of all initial spatypes.



%

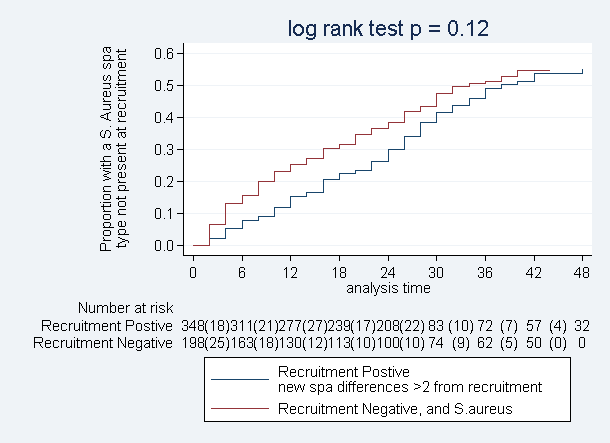


Absolute values



% restricted to 2-24 weeks

#### Gain of first new spa type: recruitment



Strate gives:

---------------------------------------------------------------------------------------------------------+

| base2 D Y Rate Lower Upper |

|--------------------------------------------------------------------------------------------------------|

| 0 126 8.1e+03 0.015497 0.013014 0.018454 |

| 1 89 4.5e+03 0.019917 0.016181 0.024516 |

Splitting at 6, 24 gives

+-------------------------------------------------------+

| new1 D Y Rate Lower Upper |

|-------------------------------------------------------|

| 0 56 3.0e+03 0.018476 0.014219 0.024008 |

| 6 105 6.7e+03 0.015593 0.012878 0.018879 |

| 24 54 2.8e+03 0.019054 0.014594 0.024879 |

+-------------------------------------------------------+

combining

+----------------------------------------------------------+

| base2 new1 D Y Rate Lower Upper |

|----------------------------------------------------------|

| 0 0 26 1.9645 13.2349 9.0113 19.4382 |

| 0 6 67 4.5480 14.7318 11.5948 18.7174 |

| 0 24 33 1.6180 20.3956 14.4997 28.6887 |

|----------------------------------------------------------|

| 1 0 30 1.0665 28.1294 19.6677 40.2316 |

| 1 6 38 2.1860 17.3833 12.6488 23.8900 |

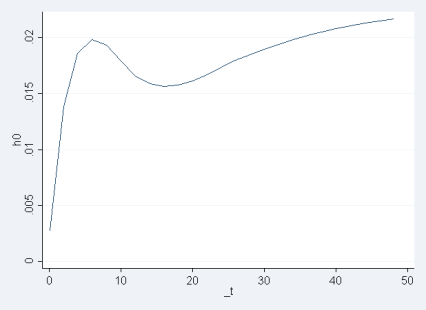
| 1 24 21 1.2160 17.2697 11.2600 26.4870 |

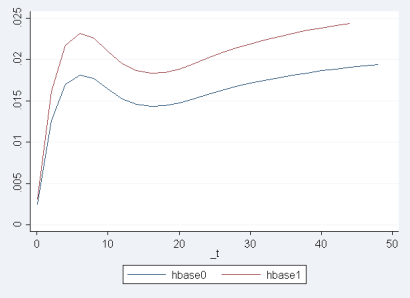
+----------------------------------------------------------+

Playing with poisson – only variable that is significant reduction in AIC = age

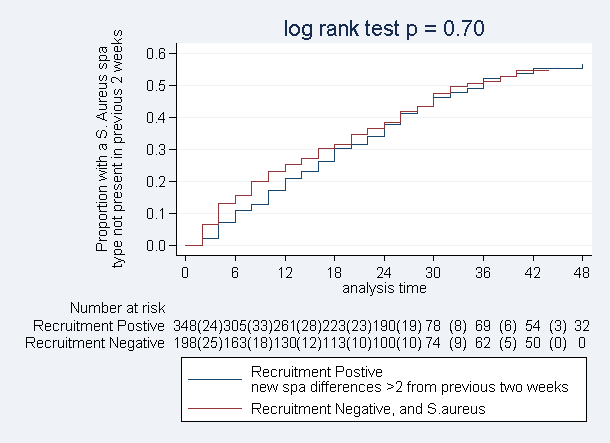
Playing with stpm2 -

Fitting hazard function to all data (sensible? Df?)

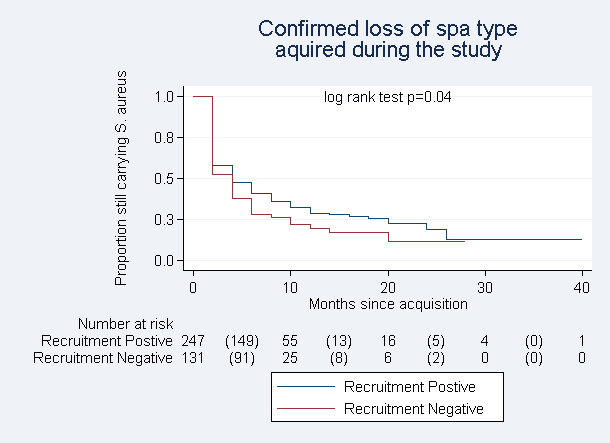




#### Gain of new spa type: previous two weeks



#### Loss of spatype acquired in study



Rate split on (6,24)

+--------------------------------------------------------+

| new1 D Y Rate Lower Upper |

|--------------------------------------------------------|

| 0 229 1.4e+03 0.163338 0.143496 0.185924 |

| 6 37 798.0000 0.046366 0.033594 0.063993 |

| 24 2 50.0000 0.040000 0.010004 0.159938 |

By base2

+----------------------------------------------------------------+

| base2 new1 D Y Rate Lower Upper |

|----------------------------------------------------------------|

| 0 0 140 938.0000 0.149254 0.126469 0.176143 |

| 0 6 25 568.0000 0.044014 0.029741 0.065138 |

| 0 24 2 46.0000 0.043478 0.010874 0.173845 |

|----------------------------------------------------------------|

| 1 0 89 464.0000 0.191810 0.155828 0.236102 |

| 1 6 12 230.0000 0.052174 0.029630 0.091870 |

| 1 24 0 4.0000 0.000000 . . |

+----------------------------------------------------------------+

Numbers too small for this split to be sensible? Try (5,10, 15) instead?

+-----------------------------------------------------------+

| base2 new1 D Y Rate (per 1000) Lower Upper |

|-----------------------------------------------------------|

| 0 0 127 0.8440 150.474 126.453 179.058 |

| 0 5 27 0.3500 77.143 52.903 112.489 |

| 0 10 6 0.1800 33.333 14.975 74.196 |

| 0 15 7 0.1780 39.326 18.748 82.490 |

| 1 0 79 0.4240 186.321 149.449 232.289 |

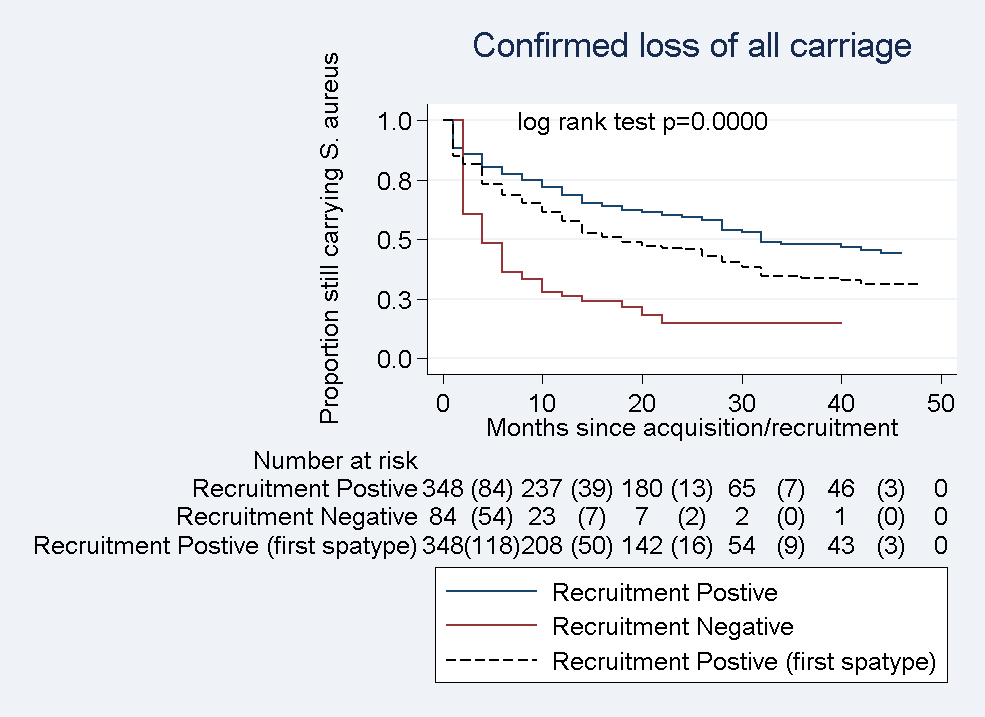
|-----------------------------------------------------------|

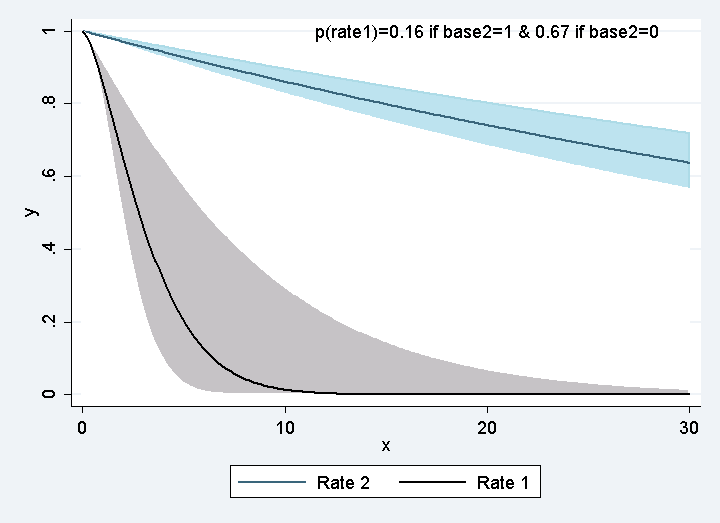
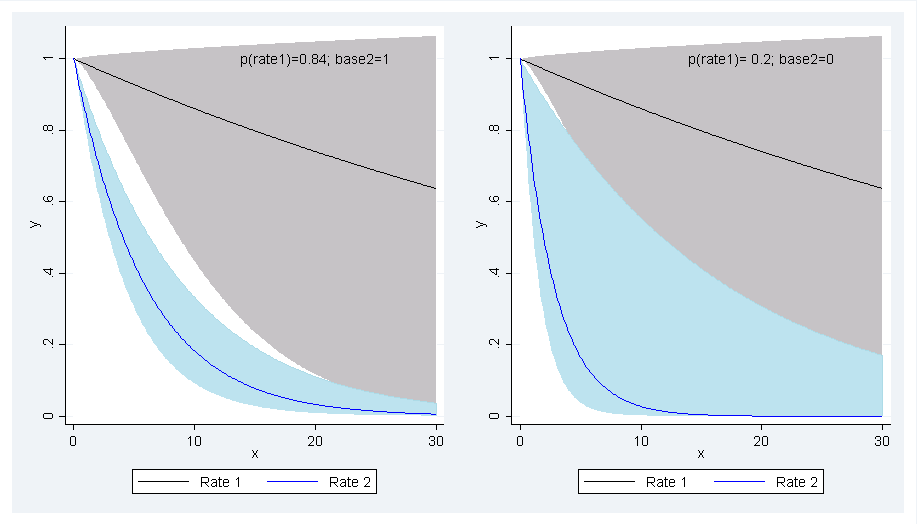
| 1 5 16 0.1440 111.111 68.070 181.367 |

| 1 10 4 0.0820 48.780 18.308 129.971 |

| 1 15 2 0.0480 41.667 10.421 166.602 |

#### Confirmed loss of all carriage



Fitting a mixture-model (weibull-exponential) to loss of all spatypes (pmix dependant on base2) Fitting a mixture-model (weibull-exponential) to the loss of all spatypes (pmix and lambda2 dependant on base2)

## Fitting Markov model to 5 state classification

Absent

One, non-persistant

Multi, non persistant

Persistant +random

persistant

Summary of attempts:

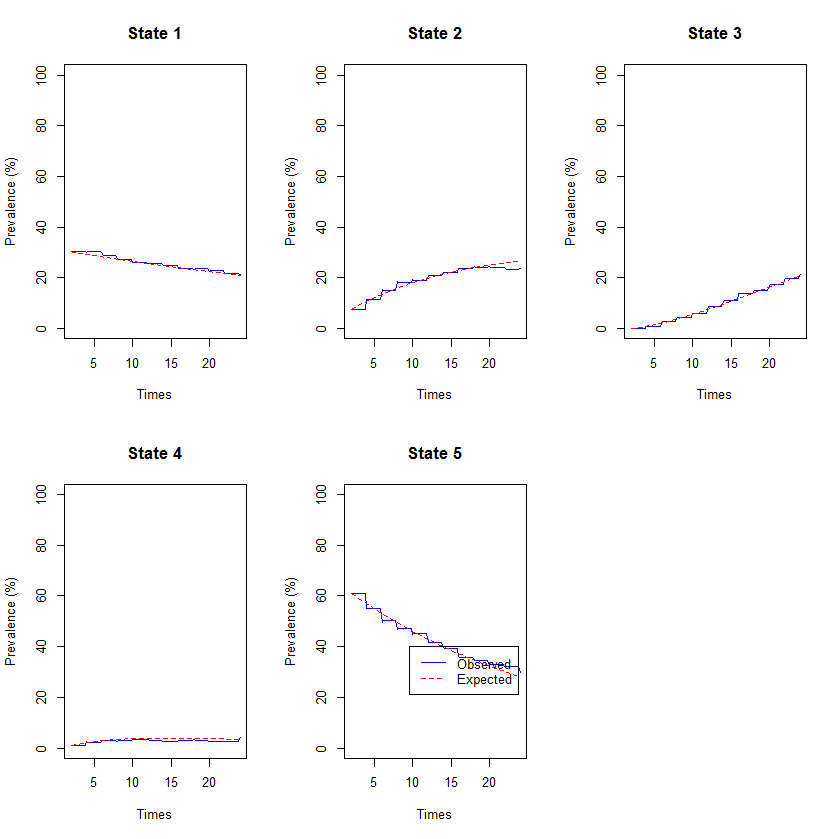
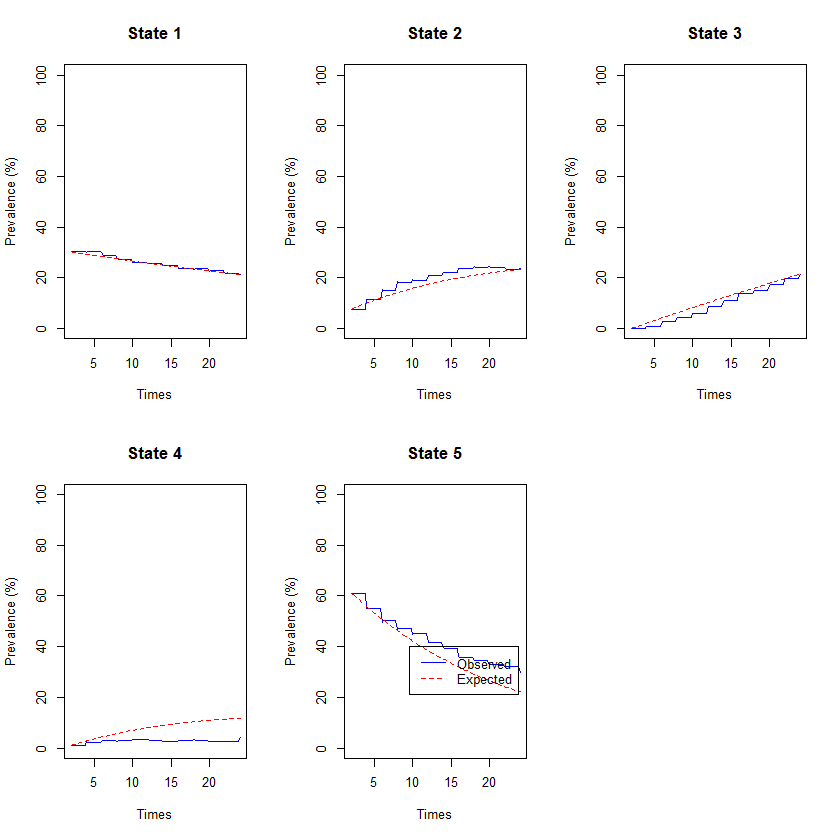
Five state, no restrictions on values -> no convergence

With restricted values -> ditto

Dropped first two measuresments -> reduces error, but still convergence issues

Dropped post 24 weeks -> still problems

Dropped error terms (note error terms tried both blanket and restricted set)  
> finally numbers start to look sensible, convergence achievable with fiddling on optim conditions. (stability not investigated yet). Dropped 1-> 3 measurement as very small number with massive error term, also cannot sensibly move 1->3 without moving through 2 at (pos. unmeasured) time.

1. Model without restrictions = fits observed data perfectly
2. 
3. Model with restrictions = slightly off
4. 

if we reintroduce error term to this model, achieves good fit with observed data BUT error terms are bizarre – decides all state 4s are misclassified and assigns them to either 3/5

Tim dislikes that there is an absorbing state/ no way back to persistence/clear. Could assign new set of states where say 4 in a row of a given spa type, or 4 clear -> back to persistence/never gain.

Staph Gain Model 12/1/2015

Considered:

"i.base2" "i.degrade" "c.age" "i.agecat" "i.age55" "i.spacat" "i.carriage" "i.ethnic" "i.male" "i.baseline\_student" "i.baseline\_iscurremp" "i.baseline\_hcrelemp" "i.baseline\_n2" "i.anti6mon" "i.antiprev"

Result:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*FINAL MODEL\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

qui streg c.age i.base2 i.antiprev, dist(weibull) anc( i.anti6mon i.spacat)

estat ic

Notes: antiprev only just significant, ethnic is significantis group as “white british” and “other”, but is that silly? Left out.

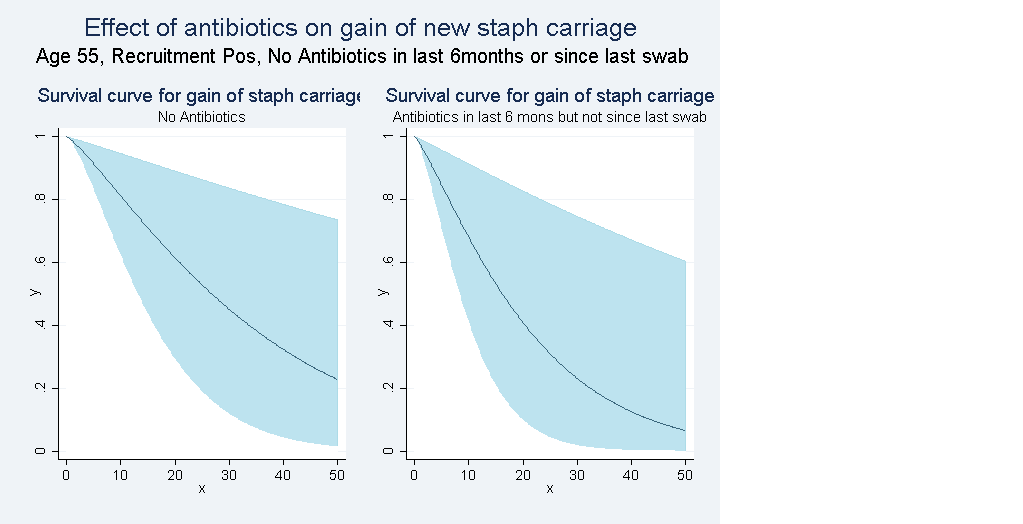
Reworked so everything is in lambda – only makes v. small change to AIC, but hazard ratios easier to explain. Dropped ethnic and antiprev, remade spacat into binary catagoty.

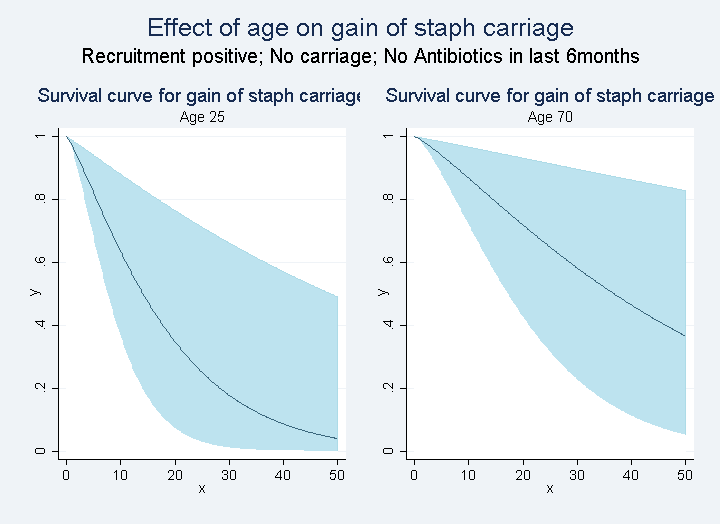
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| \_t | Coef. | [95% Conf. Interval] | Haz. Ratio | [95% Conf. Interval] |
| age | -.0258317 | ( -.0327905 , -.0188729) | .9744991 | (.9677413, .9813041) |
| 1.base2 | -.6746683 | (-1.013786, -.3355503) | .5093253 | (.3628425 , .7149446) |
| 1.anti6mon | .6097638 | ( .2216761 ,.9978516) | 1.839997 | (1.248167, 2.712448) |
| 1.carriage | -.7535823 | ( -1.096492 , -.4106722) | .4706774 | (.3340407, .6632043) |
| \_cons | -2.939879 | (-3.579583, -2.300176) |  |  |
| p | 1.214919 | (1.076596 , 1.371015) |  |  |

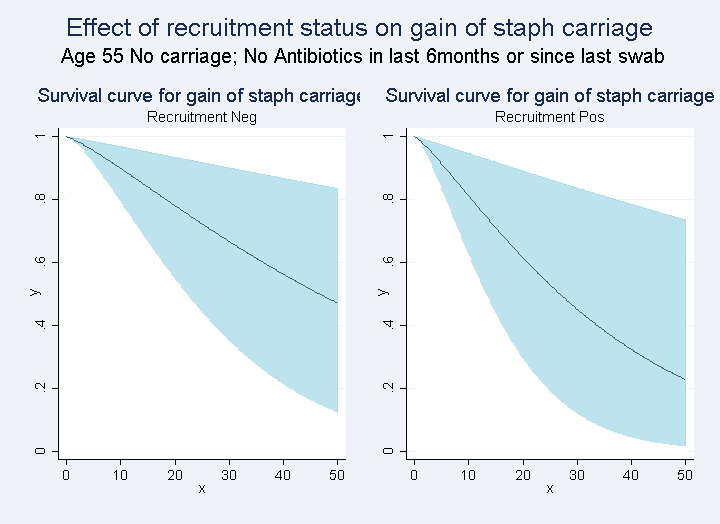
Weibull model hazard function is and survivor function is

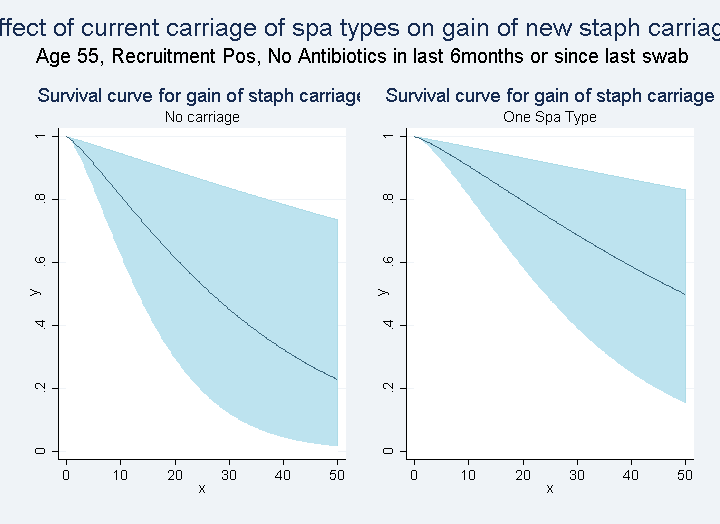
Where and p is fixed.

The hazard ratio is calculated, for example for base2=1 (which in this case means coming in to the initially negative group), by



****

****

****

## Discussion