Block-based models predicting flows into economic inactivity due to long term sickness (EILTS)

Jon Minton

Martin Taulbut

## Aim

This document will develop analyses which predict flows into EILTS using a block-based approach. This will involve developing a series of logistic regression models which predict flows into EILTS using a range of individual and household level attributes.

The underlying code begins on page 55

## Preparation

## Modelling

We will start with the simplest possible model specification, then add blocks of covariates

We are predicting whether next\_status is EILTS or not.

We will start with a manual approach to deciding on which blocks of variables, and variables within, to include

Call:  
glm(formula = becomes\_eilts ~ 1, family = binomial, data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -3.077283 0.009132 -337 <2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 102852 on 284717 degrees of freedom  
Residual deviance: 102852 on 284717 degrees of freedom  
 (62657 observations deleted due to missingness)  
AIC: 102854  
  
Number of Fisher Scoring iterations: 6

Now to add the first block of variables: history

Call:  
glm(formula = becomes\_eilts ~ this\_status, family = binomial,   
 data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -5.50920 0.03556 -154.926 <2e-16 \*\*\*  
this\_statusInactive care 2.00788 0.05577 36.000 <2e-16 \*\*\*  
this\_statusInactive long term sick 6.60620 0.04119 160.396 <2e-16 \*\*\*  
this\_statusInactive other 2.45032 0.12049 20.337 <2e-16 \*\*\*  
this\_statusInactive retired 1.94958 0.06033 32.315 <2e-16 \*\*\*  
this\_statusInactive student 0.18833 0.10198 1.847 0.0648 .   
this\_statusUnemployed 3.07013 0.04572 67.155 <2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 102852 on 284717 degrees of freedom  
Residual deviance: 44435 on 284711 degrees of freedom  
 (62657 observations deleted due to missingness)  
AIC: 44449  
  
Number of Fisher Scoring iterations: 8

Now to add the second block of variables: demographics

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 eth\_simplified, family = binomial, data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -6.63745 0.07377 -89.969 < 2e-16 \*\*\*  
this\_statusInactive care 2.10816 0.05768 36.552 < 2e-16 \*\*\*  
this\_statusInactive long term sick 6.59097 0.04198 156.988 < 2e-16 \*\*\*  
this\_statusInactive other 2.55555 0.12106 21.110 < 2e-16 \*\*\*  
this\_statusInactive retired 1.90961 0.06551 29.152 < 2e-16 \*\*\*  
this\_statusInactive student 1.03978 0.11192 9.290 < 2e-16 \*\*\*  
this\_statusUnemployed 3.24432 0.04649 69.784 < 2e-16 \*\*\*  
age\_group25-44 0.74836 0.05990 12.494 < 2e-16 \*\*\*  
age\_group45-54 1.26333 0.06075 20.797 < 2e-16 \*\*\*  
age\_group55-64 0.91921 0.06206 14.812 < 2e-16 \*\*\*  
sexmale 0.04212 0.02853 1.476 0.14   
eth\_simplifiedWhite 0.23248 0.03411 6.815 9.44e-12 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 102697 on 284023 degrees of freedom  
Residual deviance: 43766 on 284012 degrees of freedom  
 (63351 observations deleted due to missingness)  
AIC: 43790  
  
Number of Fisher Scoring iterations: 8

We can now start comparing the AICs

df AIC  
mod\_null 1 102854.06  
mod\_history 7 44448.90  
mod\_history\_demographics 12 43789.93

Now hh income, both linear and logged

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 eth\_simplified + eq\_net\_income, family = binomial, data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -6.590e+00 7.897e-02 -83.440 < 2e-16 \*\*\*  
this\_statusInactive care 2.095e+00 5.857e-02 35.768 < 2e-16 \*\*\*  
this\_statusInactive long term sick 6.571e+00 4.342e-02 151.317 < 2e-16 \*\*\*  
this\_statusInactive other 2.552e+00 1.214e-01 21.018 < 2e-16 \*\*\*  
this\_statusInactive retired 1.903e+00 6.578e-02 28.928 < 2e-16 \*\*\*  
this\_statusInactive student 1.028e+00 1.127e-01 9.124 < 2e-16 \*\*\*  
this\_statusUnemployed 3.218e+00 4.849e-02 66.366 < 2e-16 \*\*\*  
age\_group25-44 7.476e-01 6.017e-02 12.423 < 2e-16 \*\*\*  
age\_group45-54 1.266e+00 6.101e-02 20.746 < 2e-16 \*\*\*  
age\_group55-64 9.166e-01 6.233e-02 14.705 < 2e-16 \*\*\*  
sexmale 3.918e-02 2.862e-02 1.369 0.171   
eth\_simplifiedWhite 2.362e-01 3.426e-02 6.895 5.4e-12 \*\*\*  
eq\_net\_income -2.330e-05 1.445e-05 -1.613 0.107   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 102192 on 281904 degrees of freedom  
Residual deviance: 43552 on 281892 degrees of freedom  
 (65470 observations deleted due to missingness)  
AIC: 43578  
  
Number of Fisher Scoring iterations: 8

Now logged

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 eth\_simplified + log(eq\_net\_income + 0.5), family = binomial,   
 data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -6.614857 0.140640 -47.034 < 2e-16 \*\*\*  
this\_statusInactive care 2.110572 0.058341 36.176 < 2e-16 \*\*\*  
this\_statusInactive long term sick 6.587726 0.042823 153.836 < 2e-16 \*\*\*  
this\_statusInactive other 2.564306 0.121521 21.102 < 2e-16 \*\*\*  
this\_statusInactive retired 1.910969 0.065746 29.066 < 2e-16 \*\*\*  
this\_statusInactive student 1.040043 0.112704 9.228 < 2e-16 \*\*\*  
this\_statusUnemployed 3.239070 0.048420 66.895 < 2e-16 \*\*\*  
age\_group25-44 0.748574 0.060203 12.434 < 2e-16 \*\*\*  
age\_group45-54 1.265794 0.061044 20.736 < 2e-16 \*\*\*  
age\_group55-64 0.915205 0.062350 14.679 < 2e-16 \*\*\*  
sexmale 0.040290 0.028629 1.407 0.159   
eth\_simplifiedWhite 0.233709 0.034271 6.819 9.14e-12 \*\*\*  
log(eq\_net\_income + 0.5) -0.002659 0.015957 -0.167 0.868   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 102175 on 281778 degrees of freedom  
Residual deviance: 43544 on 281766 degrees of freedom  
 (65596 observations deleted due to missingness)  
AIC: 43570  
  
Number of Fisher Scoring iterations: 8

If we were to remove event history, the effect of income may be stat sig

Call:  
glm(formula = becomes\_eilts ~ age\_group + sex + eth\_simplified +   
 log(eq\_net\_income + 0.5), family = binomial, data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -1.764249 0.068014 -25.940 < 2e-16 \*\*\*  
age\_group25-44 0.987046 0.046827 21.078 < 2e-16 \*\*\*  
age\_group45-54 1.770604 0.046395 38.164 < 2e-16 \*\*\*  
age\_group55-64 1.906987 0.046450 41.054 < 2e-16 \*\*\*  
sexmale -0.001138 0.018676 -0.061 0.951   
eth\_simplifiedWhite 0.098977 0.024553 4.031 5.55e-05 \*\*\*  
log(eq\_net\_income + 0.5) -0.385585 0.007772 -49.615 < 2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 102175 on 281778 degrees of freedom  
Residual deviance: 96790 on 281772 degrees of freedom  
 (65596 observations deleted due to missingness)  
AIC: 96804  
  
Number of Fisher Scoring iterations: 7

Now car access

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 eth\_simplified + hascar, family = binomial, data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -6.23477 0.07913 -78.794 < 2e-16 \*\*\*  
this\_statusInactive care 2.04902 0.05890 34.787 < 2e-16 \*\*\*  
this\_statusInactive long term sick 6.46109 0.04336 149.023 < 2e-16 \*\*\*  
this\_statusInactive other 2.53695 0.12231 20.743 < 2e-16 \*\*\*  
this\_statusInactive retired 1.88974 0.06676 28.309 < 2e-16 \*\*\*  
this\_statusInactive student 1.00293 0.11388 8.807 < 2e-16 \*\*\*  
this\_statusUnemployed 3.08455 0.04859 63.482 < 2e-16 \*\*\*  
age\_group25-44 0.73470 0.06099 12.047 < 2e-16 \*\*\*  
age\_group45-54 1.27923 0.06187 20.675 < 2e-16 \*\*\*  
age\_group55-64 0.91998 0.06332 14.530 < 2e-16 \*\*\*  
sexmale 0.04269 0.02910 1.467 0.142   
eth\_simplifiedWhite 0.27626 0.03512 7.866 3.67e-15 \*\*\*  
hascar -0.50003 0.03074 -16.268 < 2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 99491 on 274672 degrees of freedom  
Residual deviance: 42057 on 274660 degrees of freedom  
 (72702 observations deleted due to missingness)  
AIC: 42083  
  
Number of Fisher Scoring iterations: 8

Now to add the next block of variables: health

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 eth\_simplified + hascar + lti, family = binomial, data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -6.95842 0.08619 -80.738 < 2e-16 \*\*\*  
this\_statusInactive care 1.88101 0.05956 31.582 < 2e-16 \*\*\*  
this\_statusInactive long term sick 5.50304 0.04477 122.931 < 2e-16 \*\*\*  
this\_statusInactive other 2.25620 0.12478 18.081 < 2e-16 \*\*\*  
this\_statusInactive retired 1.51623 0.06695 22.649 < 2e-16 \*\*\*  
this\_statusInactive student 1.08326 0.11507 9.414 < 2e-16 \*\*\*  
this\_statusUnemployed 2.81475 0.04962 56.727 < 2e-16 \*\*\*  
age\_group25-44 0.56288 0.06454 8.722 < 2e-16 \*\*\*  
age\_group45-54 0.95897 0.06529 14.687 < 2e-16 \*\*\*  
age\_group55-64 0.55836 0.06635 8.415 < 2e-16 \*\*\*  
sexmale 0.08040 0.02958 2.718 0.00657 \*\*   
eth\_simplifiedWhite 0.18205 0.03623 5.025 5.04e-07 \*\*\*  
hascar -0.44214 0.03110 -14.215 < 2e-16 \*\*\*  
ltiyes 2.05628 0.04236 48.545 < 2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 99400 on 274412 degrees of freedom  
Residual deviance: 39061 on 274399 degrees of freedom  
 (72962 observations deleted due to missingness)  
AIC: 39089  
  
Number of Fisher Scoring iterations: 8

Now health as continuous variables

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 hascar + eth\_simplified + mh + ph, family = binomial, data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -0.429428 0.133318 -3.221 0.00128 \*\*   
this\_statusInactive care 1.508376 0.066405 22.715 < 2e-16 \*\*\*  
this\_statusInactive long term sick 4.804996 0.051967 92.463 < 2e-16 \*\*\*  
this\_statusInactive other 2.021429 0.140777 14.359 < 2e-16 \*\*\*  
this\_statusInactive retired 1.282066 0.076053 16.857 < 2e-16 \*\*\*  
this\_statusInactive student 0.643392 0.140127 4.591 4.40e-06 \*\*\*  
this\_statusUnemployed 2.410940 0.055531 43.416 < 2e-16 \*\*\*  
age\_group25-44 0.572001 0.076125 7.514 5.73e-14 \*\*\*  
age\_group45-54 0.901074 0.077535 11.622 < 2e-16 \*\*\*  
age\_group55-64 0.451309 0.080003 5.641 1.69e-08 \*\*\*  
sexmale 0.229690 0.034090 6.738 1.61e-11 \*\*\*  
hascar -0.555532 0.035047 -15.851 < 2e-16 \*\*\*  
eth\_simplifiedWhite 0.374324 0.042506 8.806 < 2e-16 \*\*\*  
mh -0.044757 0.001315 -34.036 < 2e-16 \*\*\*  
ph -0.071427 0.001327 -53.846 < 2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 82752 on 243719 degrees of freedom  
Residual deviance: 30987 on 243705 degrees of freedom  
 (103655 observations deleted due to missingness)  
AIC: 31017  
  
Number of Fisher Scoring iterations: 8

Finally let’s look at both health and lti

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 eth\_simplified + hascar + mh + ph + lti, family = binomial,   
 data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -1.998743 0.148004 -13.505 < 2e-16 \*\*\*  
this\_statusInactive care 1.456753 0.066737 21.828 < 2e-16 \*\*\*  
this\_statusInactive long term sick 4.524810 0.052024 86.975 < 2e-16 \*\*\*  
this\_statusInactive other 1.921301 0.141816 13.548 < 2e-16 \*\*\*  
this\_statusInactive retired 1.125488 0.075513 14.905 < 2e-16 \*\*\*  
this\_statusInactive student 0.720100 0.140736 5.117 3.11e-07 \*\*\*  
this\_statusUnemployed 2.339414 0.056058 41.732 < 2e-16 \*\*\*  
age\_group25-44 0.477625 0.078062 6.119 9.44e-10 \*\*\*  
age\_group45-54 0.755825 0.079423 9.516 < 2e-16 \*\*\*  
age\_group55-64 0.304891 0.081493 3.741 0.000183 \*\*\*  
sexmale 0.212207 0.034089 6.225 4.81e-10 \*\*\*  
eth\_simplifiedWhite 0.290800 0.043041 6.756 1.42e-11 \*\*\*  
hascar -0.522201 0.035132 -14.864 < 2e-16 \*\*\*  
mh -0.036517 0.001341 -27.236 < 2e-16 \*\*\*  
ph -0.057356 0.001408 -40.739 < 2e-16 \*\*\*  
ltiyes 1.297349 0.051534 25.174 < 2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 82687 on 243549 degrees of freedom  
Residual deviance: 30229 on 243534 degrees of freedom  
 (103825 observations deleted due to missingness)  
AIC: 30261  
  
Number of Fisher Scoring iterations: 8

But is the AIC of including both types of health better than just sf12 or lti?

df AIC  
mod\_history\_demographics\_car\_sf12 15 31016.72  
mod\_history\_demographics\_car\_lti 14 39089.14  
mod\_history\_demographics\_car\_health 16 30260.91

The number of observations aren’t exactly the same, so we can’t directly compare AICs. However, the AIC of the model with both types of health is lower than the AIC of the model with just one type of health. However it appears that the sf12 derived variables are more useful than the lti/no lti binary variable

So let’s use mh and ph, before moving onto the next block of variables: household

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 eth\_simplified + lti + mh + ph + has\_children, family = binomial,   
 data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -2.126822 0.145400 -14.627 < 2e-16 \*\*\*  
this\_statusInactive care 1.616602 0.066401 24.346 < 2e-16 \*\*\*  
this\_statusInactive long term sick 4.633204 0.050543 91.668 < 2e-16 \*\*\*  
this\_statusInactive other 1.951821 0.140017 13.940 < 2e-16 \*\*\*  
this\_statusInactive retired 1.139327 0.073956 15.406 < 2e-16 \*\*\*  
this\_statusInactive student 0.808770 0.139328 5.805 6.45e-09 \*\*\*  
this\_statusUnemployed 2.491928 0.053899 46.234 < 2e-16 \*\*\*  
age\_group25-44 0.515059 0.077150 6.676 2.45e-11 \*\*\*  
age\_group45-54 0.676430 0.078712 8.594 < 2e-16 \*\*\*  
age\_group55-64 0.148588 0.081717 1.818 0.069 .   
sexmale 0.192425 0.033656 5.717 1.08e-08 \*\*\*  
eth\_simplifiedWhite 0.212236 0.042596 4.983 6.28e-07 \*\*\*  
ltiyes 1.294258 0.050967 25.394 < 2e-16 \*\*\*  
mh -0.037514 0.001322 -28.379 < 2e-16 \*\*\*  
ph -0.057109 0.001387 -41.188 < 2e-16 \*\*\*  
has\_childrenTRUE -0.402256 0.038110 -10.555 < 2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 84625 on 248916 degrees of freedom  
Residual deviance: 31047 on 248901 degrees of freedom  
 (98458 observations deleted due to missingness)  
AIC: 31079  
  
Number of Fisher Scoring iterations: 8

Now hh category

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 hascar + eth\_simplified + lti + mh + ph + hh\_type, family = binomial,   
 data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -1.444565 0.203602 -7.095 1.29e-12 \*\*\*  
this\_statusInactive care 1.486190 0.084503 17.588 < 2e-16 \*\*\*  
this\_statusInactive long term sick 4.395104 0.063427 69.294 < 2e-16 \*\*\*  
this\_statusInactive other 1.957847 0.173866 11.261 < 2e-16 \*\*\*  
this\_statusInactive retired 1.229633 0.091304 13.467 < 2e-16 \*\*\*  
this\_statusInactive student 0.928661 0.189587 4.898 9.67e-07 \*\*\*  
this\_statusUnemployed 2.256186 0.069163 32.621 < 2e-16 \*\*\*  
age\_group25-44 0.678116 0.125470 5.405 6.49e-08 \*\*\*  
age\_group45-54 0.913669 0.128504 7.110 1.16e-12 \*\*\*  
age\_group55-64 0.572838 0.133064 4.305 1.67e-05 \*\*\*  
sexmale 0.093554 0.043920 2.130 0.0332 \*   
hascar -0.486297 0.047316 -10.278 < 2e-16 \*\*\*  
eth\_simplifiedWhite 0.224398 0.056841 3.948 7.89e-05 \*\*\*  
ltiyes 1.235665 0.063929 19.329 < 2e-16 \*\*\*  
mh -0.038245 0.001652 -23.150 < 2e-16 \*\*\*  
ph -0.060273 0.001723 -34.989 < 2e-16 \*\*\*  
hh\_typeSmall Adult -0.376975 0.059776 -6.306 2.85e-10 \*\*\*  
hh\_typeSingle Parent -0.429186 0.077776 -5.518 3.42e-08 \*\*\*  
hh\_typeFamily with 1-2 Children -0.596903 0.067558 -8.835 < 2e-16 \*\*\*  
hh\_typeFamily with 3 or more Children -0.775609 0.095714 -8.103 5.34e-16 \*\*\*  
hh\_typeSingle Pensioner -1.074282 0.116606 -9.213 < 2e-16 \*\*\*  
hh\_typePensioner Couple -0.894505 0.081699 -10.949 < 2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 55849 on 166110 degrees of freedom  
Residual deviance: 20011 on 166089 degrees of freedom  
 (181264 observations deleted due to missingness)  
AIC: 20055  
  
Number of Fisher Scoring iterations: 8

Now qualifications, which we think is the last block

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 hascar + eth\_simplified + lti + mh + ph + hh\_type + hiqual\_dv,   
 family = binomial, data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -1.489855 0.209272 -7.119 1.09e-12 \*\*\*  
this\_statusInactive care 1.383321 0.085565 16.167 < 2e-16 \*\*\*  
this\_statusInactive long term sick 4.292269 0.064163 66.896 < 2e-16 \*\*\*  
this\_statusInactive other 1.907683 0.174489 10.933 < 2e-16 \*\*\*  
this\_statusInactive retired 1.188993 0.091991 12.925 < 2e-16 \*\*\*  
this\_statusInactive student 0.911707 0.191371 4.764 1.90e-06 \*\*\*  
this\_statusUnemployed 2.159671 0.070122 30.799 < 2e-16 \*\*\*  
age\_group25-44 0.714528 0.127378 5.609 2.03e-08 \*\*\*  
age\_group45-54 0.922738 0.130664 7.062 1.64e-12 \*\*\*  
age\_group55-64 0.545290 0.135435 4.026 5.67e-05 \*\*\*  
sexmale 0.076076 0.044292 1.718 0.08587 .   
hascar -0.394998 0.048476 -8.148 3.69e-16 \*\*\*  
eth\_simplifiedWhite 0.181965 0.057622 3.158 0.00159 \*\*   
ltiyes 1.227335 0.064174 19.125 < 2e-16 \*\*\*  
mh -0.037589 0.001659 -22.662 < 2e-16 \*\*\*  
ph -0.059200 0.001737 -34.079 < 2e-16 \*\*\*  
hh\_typeSmall Adult -0.400592 0.060216 -6.653 2.88e-11 \*\*\*  
hh\_typeSingle Parent -0.468823 0.078287 -5.989 2.12e-09 \*\*\*  
hh\_typeFamily with 1-2 Children -0.605993 0.067919 -8.922 < 2e-16 \*\*\*  
hh\_typeFamily with 3 or more Children -0.824808 0.096061 -8.586 < 2e-16 \*\*\*  
hh\_typeSingle Pensioner -1.125300 0.116953 -9.622 < 2e-16 \*\*\*  
hh\_typePensioner Couple -0.951806 0.082293 -11.566 < 2e-16 \*\*\*  
hiqual\_dvDegree -0.336396 0.073994 -4.546 5.46e-06 \*\*\*  
hiqual\_dvGCSE etc 0.111135 0.061660 1.802 0.07149 .   
hiqual\_dvNo qualification 0.342377 0.066033 5.185 2.16e-07 \*\*\*  
hiqual\_dvOther higher degree -0.137755 0.080659 -1.708 0.08766 .   
hiqual\_dvOther qualification 0.198317 0.071517 2.773 0.00555 \*\*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 55525 on 164794 degrees of freedom  
Residual deviance: 19821 on 164768 degrees of freedom  
 (182580 observations deleted due to missingness)  
AIC: 19875  
  
Number of Fisher Scoring iterations: 8

We now have a series of models, organised into blocks, which build up in complexity incrementally. Each appears to increase the proportion explained. Let’s use Nagelkerke’s R^2 to compare the models

$N  
[1] 284718  
  
$R2  
[1] 5.303035e-12

$N  
[1] 284718  
  
$R2  
[1] 0.6118141

$N  
[1] 284024  
  
$R2  
[1] 0.6175377

$N  
[1] 274673  
  
$R2  
[1] 0.6209505

$N  
[1] 243550  
  
$R2  
[1] 0.6731046

$N  
[1] 248917  
  
$R2  
[1] 0.6719252

$N  
[1] 166111  
  
$R2  
[1] 0.6796468

$N  
[1] 164795  
  
$R2  
[1] 0.6810013

Now to make this a table

| label | nr2 |
| --- | --- |
| null | 0.0000000 |
| history | 0.6118141 |
| history\_demographics | 0.6175377 |
| history\_demographics\_hascar | 0.6209505 |
| history\_demographics\_car\_health | 0.6731046 |
| history\_demographics\_health\_car\_hhchildren | 0.6719252 |
| history\_demographics\_health\_car\_hhtype | 0.6796468 |
| history\_demographics\_health\_car\_hhtype\_qual | 0.6810013 |

Let’s return to the demographics block and see if we can do better (using the spec previously arrived at )

Call:  
glm(formula = becomes\_eilts ~ this\_status \* sex + splines::bs(age,   
 5) + sex + eth\_simplified, family = binomial, data = data\_tidied)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -7.29911 0.15367 -47.497 < 2e-16  
this\_statusInactive care 2.00969 0.06578 30.553 < 2e-16  
this\_statusInactive long term sick 6.48097 0.05583 116.075 < 2e-16  
this\_statusInactive other 2.30230 0.16892 13.630 < 2e-16  
this\_statusInactive retired 1.98284 0.08890 22.305 < 2e-16  
this\_statusInactive student 1.18139 0.15326 7.709 1.27e-14  
this\_statusUnemployed 3.20384 0.06413 49.959 < 2e-16  
sexmale -0.17094 0.07176 -2.382 0.017213  
splines::bs(age, 5)1 1.07505 0.22816 4.712 2.45e-06  
splines::bs(age, 5)2 1.39106 0.14643 9.500 < 2e-16  
splines::bs(age, 5)3 1.84290 0.18022 10.226 < 2e-16  
splines::bs(age, 5)4 2.53725 0.15095 16.809 < 2e-16  
splines::bs(age, 5)5 0.47966 0.15986 3.000 0.002695  
eth\_simplifiedWhite 0.23946 0.03439 6.964 3.31e-12  
this\_statusInactive care:sexmale 0.31402 0.18442 1.703 0.088620  
this\_statusInactive long term sick:sexmale 0.31299 0.08365 3.742 0.000183  
this\_statusInactive other:sexmale 0.56931 0.24269 2.346 0.018985  
this\_statusInactive retired:sexmale 0.62219 0.12161 5.116 3.11e-07  
this\_statusInactive student:sexmale 0.39506 0.20453 1.932 0.053418  
this\_statusUnemployed:sexmale 0.15391 0.09239 1.666 0.095743  
   
(Intercept) \*\*\*  
this\_statusInactive care \*\*\*  
this\_statusInactive long term sick \*\*\*  
this\_statusInactive other \*\*\*  
this\_statusInactive retired \*\*\*  
this\_statusInactive student \*\*\*  
this\_statusUnemployed \*\*\*  
sexmale \*   
splines::bs(age, 5)1 \*\*\*  
splines::bs(age, 5)2 \*\*\*  
splines::bs(age, 5)3 \*\*\*  
splines::bs(age, 5)4 \*\*\*  
splines::bs(age, 5)5 \*\*   
eth\_simplifiedWhite \*\*\*  
this\_statusInactive care:sexmale .   
this\_statusInactive long term sick:sexmale \*\*\*  
this\_statusInactive other:sexmale \*   
this\_statusInactive retired:sexmale \*\*\*  
this\_statusInactive student:sexmale .   
this\_statusUnemployed:sexmale .   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 102697 on 284023 degrees of freedom  
Residual deviance: 43199 on 284004 degrees of freedom  
 (63351 observations deleted due to missingness)  
AIC: 43239  
  
Number of Fisher Scoring iterations: 8

$N  
[1] 284024  
  
$R2  
[1] 0.6228821

Let’s now look at the stepwise AIC approach to see if similar variables are selected

Start: AIC=23518.89  
becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified  
  
 Df Deviance AIC  
+ ph 1 21774 21800  
+ lti 1 21851 21877  
+ mh 1 22825 22851  
+ hh\_type 6 23117 23153  
+ hiqual\_dv 5 23225 23259  
+ hascar 1 23244 23270  
- sex 1 23495 23517  
<none> 23495 23519  
- eth\_simplified 1 23500 23522  
- age\_group 3 23764 23782  
- this\_status 6 53511 53523  
  
Step: AIC=21800.33  
becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified +   
 ph  
  
 Df Deviance AIC  
+ mh 1 20839 20867  
+ lti 1 20977 21005  
+ hh\_type 6 21273 21311  
+ hascar 1 21436 21464  
+ hiqual\_dv 5 21591 21627  
<none> 21774 21800  
- sex 1 21779 21803  
- eth\_simplified 1 21785 21809  
- age\_group 3 21949 21969  
- ph 1 23495 23519  
- this\_status 6 37059 37073  
  
Step: AIC=20866.64  
becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified +   
 ph + mh  
  
 Df Deviance AIC  
+ lti 1 20367 20397  
+ hh\_type 6 20440 20480  
+ hascar 1 20577 20607  
+ hiqual\_dv 5 20695 20733  
<none> 20839 20867  
- eth\_simplified 1 20851 20877  
- sex 1 20868 20894  
- age\_group 3 20968 20990  
- mh 1 21774 21800  
- ph 1 22825 22851  
- this\_status 6 31449 31465  
  
Step: AIC=20396.82  
becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified +   
 ph + mh + lti  
  
 Df Deviance AIC  
+ hh\_type 6 20024 20066  
+ hascar 1 20138 20170  
+ hiqual\_dv 5 20231 20271  
<none> 20367 20397  
- eth\_simplified 1 20372 20400  
- sex 1 20392 20420  
- age\_group 3 20471 20495  
- lti 1 20839 20867  
- mh 1 20977 21005  
- ph 1 21497 21525  
- this\_status 6 29985 30003  
  
Step: AIC=20065.83  
becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified +   
 ph + mh + lti + hh\_type  
  
 Df Deviance AIC  
+ hiqual\_dv 5 19886 19938  
+ hascar 1 19920 19964  
<none> 20024 20066  
- sex 1 20030 20070  
- eth\_simplified 1 20033 20073  
- age\_group 3 20093 20129  
- hh\_type 6 20367 20397  
- lti 1 20440 20480  
- mh 1 20574 20614  
- ph 1 21243 21283  
- this\_status 6 28682 28712  
  
Step: AIC=19938.15  
becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified +   
 ph + mh + lti + hh\_type + hiqual\_dv  
  
 Df Deviance AIC  
+ hascar 1 19820 19874  
<none> 19886 19938  
- sex 1 19890 19940  
- eth\_simplified 1 19892 19942  
- age\_group 3 19966 20012  
- hiqual\_dv 5 20024 20066  
- hh\_type 6 20231 20271  
- lti 1 20295 20345  
- mh 1 20411 20461  
- ph 1 21045 21095  
- this\_status 6 27646 27686  
  
Step: AIC=19874.06  
becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified +   
 ph + mh + lti + hh\_type + hiqual\_dv + hascar  
  
 Df Deviance AIC  
<none> 19820 19874  
- sex 1 19823 19875  
- eth\_simplified 1 19830 19882  
- hascar 1 19886 19938  
- age\_group 3 19903 19951  
- hiqual\_dv 5 19920 19964  
- hh\_type 6 20059 20101  
- lti 1 20221 20273  
- mh 1 20335 20387  
- ph 1 21011 21063  
- this\_status 6 27063 27105

Let’s see the final specification arrived at

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 eth\_simplified + ph + mh + lti + hh\_type + hiqual\_dv + hascar,   
 family = binomial, data = complete\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -1.489860 0.209268 -7.119 1.08e-12 \*\*\*  
this\_statusInactive care 1.383259 0.085565 16.166 < 2e-16 \*\*\*  
this\_statusInactive long term sick 4.292131 0.064162 66.895 < 2e-16 \*\*\*  
this\_statusInactive other 1.909042 0.174523 10.939 < 2e-16 \*\*\*  
this\_statusInactive retired 1.189942 0.091990 12.936 < 2e-16 \*\*\*  
this\_statusInactive student 0.911692 0.191369 4.764 1.90e-06 \*\*\*  
this\_statusUnemployed 2.159479 0.070123 30.796 < 2e-16 \*\*\*  
age\_group25-44 0.714655 0.127377 5.611 2.02e-08 \*\*\*  
age\_group45-54 0.922760 0.130662 7.062 1.64e-12 \*\*\*  
age\_group55-64 0.545435 0.135433 4.027 5.64e-05 \*\*\*  
sexmale 0.076283 0.044294 1.722 0.08503 .   
eth\_simplifiedWhite 0.182040 0.057622 3.159 0.00158 \*\*   
ph -0.059198 0.001737 -34.079 < 2e-16 \*\*\*  
mh -0.037595 0.001659 -22.666 < 2e-16 \*\*\*  
ltiyes 1.227407 0.064173 19.126 < 2e-16 \*\*\*  
hh\_typeSmall Adult -0.399958 0.060224 -6.641 3.11e-11 \*\*\*  
hh\_typeSingle Parent -0.468642 0.078288 -5.986 2.15e-09 \*\*\*  
hh\_typeFamily with 1-2 Children -0.605696 0.067922 -8.918 < 2e-16 \*\*\*  
hh\_typeFamily with 3 or more Children -0.824577 0.096063 -8.584 < 2e-16 \*\*\*  
hh\_typeSingle Pensioner -1.125580 0.116945 -9.625 < 2e-16 \*\*\*  
hh\_typePensioner Couple -0.951679 0.082292 -11.565 < 2e-16 \*\*\*  
hiqual\_dvDegree -0.336249 0.073996 -4.544 5.52e-06 \*\*\*  
hiqual\_dvGCSE etc 0.111043 0.061661 1.801 0.07173 .   
hiqual\_dvNo qualification 0.342532 0.066037 5.187 2.14e-07 \*\*\*  
hiqual\_dvOther higher degree -0.137846 0.080658 -1.709 0.08745 .   
hiqual\_dvOther qualification 0.198179 0.071517 2.771 0.00559 \*\*   
hascar -0.395392 0.048482 -8.155 3.48e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 55524 on 164773 degrees of freedom  
Residual deviance: 19820 on 164747 degrees of freedom  
AIC: 19874  
  
Number of Fisher Scoring iterations: 8

Let’s see what happens if we ask the algorithm to prune our final model specification

Start: AIC=19874.06  
becomes\_eilts ~ this\_status + age\_group + sex + hascar + eth\_simplified +   
 lti + mh + ph + hh\_type + hiqual\_dv  
  
 Df Deviance AIC  
<none> 19820 19874  
- sex 1 19823 19875  
- eth\_simplified 1 19830 19882  
- hascar 1 19886 19938  
- age\_group 3 19903 19951  
- hiqual\_dv 5 19920 19964  
- hh\_type 6 20059 20101  
- lti 1 20221 20273  
- mh 1 20335 20387  
- ph 1 21011 21063  
- this\_status 6 27063 27105

Call:  
glm(formula = becomes\_eilts ~ this\_status + age\_group + sex +   
 hascar + eth\_simplified + lti + mh + ph + hh\_type + hiqual\_dv,   
 family = binomial, data = complete\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -1.489860 0.209268 -7.119 1.08e-12 \*\*\*  
this\_statusInactive care 1.383259 0.085565 16.166 < 2e-16 \*\*\*  
this\_statusInactive long term sick 4.292131 0.064162 66.895 < 2e-16 \*\*\*  
this\_statusInactive other 1.909042 0.174523 10.939 < 2e-16 \*\*\*  
this\_statusInactive retired 1.189942 0.091990 12.936 < 2e-16 \*\*\*  
this\_statusInactive student 0.911692 0.191369 4.764 1.90e-06 \*\*\*  
this\_statusUnemployed 2.159479 0.070123 30.796 < 2e-16 \*\*\*  
age\_group25-44 0.714655 0.127377 5.611 2.02e-08 \*\*\*  
age\_group45-54 0.922760 0.130662 7.062 1.64e-12 \*\*\*  
age\_group55-64 0.545435 0.135433 4.027 5.64e-05 \*\*\*  
sexmale 0.076283 0.044294 1.722 0.08503 .   
hascar -0.395392 0.048482 -8.155 3.48e-16 \*\*\*  
eth\_simplifiedWhite 0.182040 0.057622 3.159 0.00158 \*\*   
ltiyes 1.227407 0.064173 19.126 < 2e-16 \*\*\*  
mh -0.037595 0.001659 -22.666 < 2e-16 \*\*\*  
ph -0.059198 0.001737 -34.079 < 2e-16 \*\*\*  
hh\_typeSmall Adult -0.399958 0.060224 -6.641 3.11e-11 \*\*\*  
hh\_typeSingle Parent -0.468642 0.078288 -5.986 2.15e-09 \*\*\*  
hh\_typeFamily with 1-2 Children -0.605696 0.067922 -8.918 < 2e-16 \*\*\*  
hh\_typeFamily with 3 or more Children -0.824577 0.096063 -8.584 < 2e-16 \*\*\*  
hh\_typeSingle Pensioner -1.125580 0.116945 -9.625 < 2e-16 \*\*\*  
hh\_typePensioner Couple -0.951679 0.082292 -11.565 < 2e-16 \*\*\*  
hiqual\_dvDegree -0.336249 0.073996 -4.544 5.52e-06 \*\*\*  
hiqual\_dvGCSE etc 0.111043 0.061661 1.801 0.07173 .   
hiqual\_dvNo qualification 0.342532 0.066037 5.187 2.14e-07 \*\*\*  
hiqual\_dvOther higher degree -0.137846 0.080658 -1.709 0.08745 .   
hiqual\_dvOther qualification 0.198179 0.071517 2.771 0.00559 \*\*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 55524 on 164773 degrees of freedom  
Residual deviance: 19820 on 164747 degrees of freedom  
AIC: 19874  
  
Number of Fisher Scoring iterations: 8

The pruned model is the same as the model we derived manually, so all variables appear essential

Finally, we can start with the simplest and move forwards

Start: AIC=55525.76  
becomes\_eilts ~ 1  
  
 Df Deviance AIC  
+ this\_status 6 23771 23785  
+ ph 1 37494 37498  
+ lti 1 42552 42556  
+ mh 1 48351 48355  
+ hiqual\_dv 5 50580 50592  
+ hascar 1 51287 51291  
+ hh\_type 6 51774 51788  
+ age\_group 3 53516 53524  
+ eth\_simplified 1 55509 55513  
<none> 55524 55526  
+ sex 1 55522 55526  
  
Step: AIC=23784.67  
becomes\_eilts ~ this\_status  
  
 Df Deviance AIC  
+ ph 1 21961 21977  
+ lti 1 22004 22020  
+ mh 1 23139 23155  
+ hh\_type 6 23331 23357  
+ hiqual\_dv 5 23494 23518  
+ age\_group 3 23500 23520  
+ hascar 1 23563 23579  
+ eth\_simplified 1 23764 23780  
<none> 23771 23785  
+ sex 1 23770 23786  
  
Step: AIC=21976.79  
becomes\_eilts ~ this\_status + ph  
  
 Df Deviance AIC  
+ mh 1 21011 21029  
+ lti 1 21120 21138  
+ hh\_type 6 21411 21439  
+ hascar 1 21653 21671  
+ age\_group 3 21790 21812  
+ hiqual\_dv 5 21795 21821  
+ eth\_simplified 1 21953 21971  
+ sex 1 21956 21974  
<none> 21961 21977  
  
Step: AIC=21029.14  
becomes\_eilts ~ this\_status + ph + mh  
  
 Df Deviance AIC  
+ lti 1 20500 20520  
+ hh\_type 6 20549 20579  
+ hascar 1 20778 20798  
+ hiqual\_dv 5 20872 20900  
+ age\_group 3 20881 20905  
+ sex 1 20979 20999  
+ eth\_simplified 1 21000 21020  
<none> 21011 21029  
  
Step: AIC=20520.09  
becomes\_eilts ~ this\_status + ph + mh + lti  
  
 Df Deviance AIC  
+ hh\_type 6 20105 20137  
+ hascar 1 20292 20314  
+ hiqual\_dv 5 20374 20404  
+ age\_group 3 20397 20423  
+ sex 1 20476 20498  
+ eth\_simplified 1 20496 20518  
<none> 20500 20520  
  
Step: AIC=20137.17  
becomes\_eilts ~ this\_status + ph + mh + lti + hh\_type  
  
 Df Deviance AIC  
+ hiqual\_dv 5 19973 20015  
+ hascar 1 20014 20048  
+ age\_group 3 20039 20077  
+ eth\_simplified 1 20097 20131  
+ sex 1 20101 20135  
<none> 20105 20137  
  
Step: AIC=20014.87  
becomes\_eilts ~ this\_status + ph + mh + lti + hh\_type + hiqual\_dv  
  
 Df Deviance AIC  
+ age\_group 3 19896 19944  
+ hascar 1 19914 19958  
+ eth\_simplified 1 19968 20012  
+ sex 1 19971 20015  
<none> 19973 20015  
  
Step: AIC=19943.58  
becomes\_eilts ~ this\_status + ph + mh + lti + hh\_type + hiqual\_dv +   
 age\_group  
  
 Df Deviance AIC  
+ hascar 1 19833 19883  
+ eth\_simplified 1 19890 19940  
+ sex 1 19892 19942  
<none> 19896 19944  
  
Step: AIC=19883.08  
becomes\_eilts ~ this\_status + ph + mh + lti + hh\_type + hiqual\_dv +   
 age\_group + hascar  
  
 Df Deviance AIC  
+ eth\_simplified 1 19823 19875  
+ sex 1 19830 19882  
<none> 19833 19883  
  
Step: AIC=19875.02  
becomes\_eilts ~ this\_status + ph + mh + lti + hh\_type + hiqual\_dv +   
 age\_group + hascar + eth\_simplified  
  
 Df Deviance AIC  
+ sex 1 19820 19874  
<none> 19823 19875  
  
Step: AIC=19874.06  
becomes\_eilts ~ this\_status + ph + mh + lti + hh\_type + hiqual\_dv +   
 age\_group + hascar + eth\_simplified + sex

Call:  
glm(formula = becomes\_eilts ~ this\_status + ph + mh + lti + hh\_type +   
 hiqual\_dv + age\_group + hascar + eth\_simplified + sex, family = binomial,   
 data = complete\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -1.489860 0.209268 -7.119 1.08e-12 \*\*\*  
this\_statusInactive care 1.383259 0.085565 16.166 < 2e-16 \*\*\*  
this\_statusInactive long term sick 4.292131 0.064162 66.895 < 2e-16 \*\*\*  
this\_statusInactive other 1.909042 0.174523 10.939 < 2e-16 \*\*\*  
this\_statusInactive retired 1.189942 0.091990 12.936 < 2e-16 \*\*\*  
this\_statusInactive student 0.911692 0.191369 4.764 1.90e-06 \*\*\*  
this\_statusUnemployed 2.159479 0.070123 30.796 < 2e-16 \*\*\*  
ph -0.059198 0.001737 -34.079 < 2e-16 \*\*\*  
mh -0.037595 0.001659 -22.666 < 2e-16 \*\*\*  
ltiyes 1.227407 0.064173 19.126 < 2e-16 \*\*\*  
hh\_typeSmall Adult -0.399958 0.060224 -6.641 3.11e-11 \*\*\*  
hh\_typeSingle Parent -0.468642 0.078288 -5.986 2.15e-09 \*\*\*  
hh\_typeFamily with 1-2 Children -0.605696 0.067922 -8.918 < 2e-16 \*\*\*  
hh\_typeFamily with 3 or more Children -0.824577 0.096063 -8.584 < 2e-16 \*\*\*  
hh\_typeSingle Pensioner -1.125580 0.116945 -9.625 < 2e-16 \*\*\*  
hh\_typePensioner Couple -0.951679 0.082292 -11.565 < 2e-16 \*\*\*  
hiqual\_dvDegree -0.336249 0.073996 -4.544 5.52e-06 \*\*\*  
hiqual\_dvGCSE etc 0.111043 0.061661 1.801 0.07173 .   
hiqual\_dvNo qualification 0.342532 0.066037 5.187 2.14e-07 \*\*\*  
hiqual\_dvOther higher degree -0.137846 0.080658 -1.709 0.08745 .   
hiqual\_dvOther qualification 0.198179 0.071517 2.771 0.00559 \*\*   
age\_group25-44 0.714655 0.127377 5.611 2.02e-08 \*\*\*  
age\_group45-54 0.922760 0.130662 7.062 1.64e-12 \*\*\*  
age\_group55-64 0.545435 0.135433 4.027 5.64e-05 \*\*\*  
hascar -0.395392 0.048482 -8.155 3.48e-16 \*\*\*  
eth\_simplifiedWhite 0.182040 0.057622 3.159 0.00158 \*\*   
sexmale 0.076283 0.044294 1.722 0.08503 .   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 55524 on 164773 degrees of freedom  
Residual deviance: 19820 on 164747 degrees of freedom  
AIC: 19874  
  
Number of Fisher Scoring iterations: 8

Once again we end up with the same model as we derived manually, so all variables appear essential.

Finally, for now, we’ll look at just those who start off in the unemployed category. This means the history variable is no longer needed as everyone’s history is now the same

Start: AIC=3454.01  
becomes\_eilts ~ age\_group + sex + hascar + eth\_simplified + lti +   
 mh + ph + hh\_type + hiqual\_dv + eq\_net\_income  
  
 Df Deviance AIC  
- sex 1 3410.2 3452.2  
- eth\_simplified 1 3410.6 3452.6  
<none> 3410.0 3454.0  
- eq\_net\_income 1 3418.9 3460.9  
- hascar 1 3422.5 3464.5  
- age\_group 3 3427.3 3465.3  
- hiqual\_dv 5 3449.8 3483.8  
- hh\_type 6 3464.4 3496.4  
- lti 1 3505.3 3547.3  
- mh 1 3533.7 3575.7  
- ph 1 3624.5 3666.5  
  
Step: AIC=3452.24  
becomes\_eilts ~ age\_group + hascar + eth\_simplified + lti + mh +   
 ph + hh\_type + hiqual\_dv + eq\_net\_income  
  
 Df Deviance AIC  
- eth\_simplified 1 3410.8 3450.8  
<none> 3410.2 3452.2  
- eq\_net\_income 1 3419.0 3459.0  
- hascar 1 3422.8 3462.8  
- age\_group 3 3427.6 3463.6  
- hiqual\_dv 5 3450.0 3482.0  
- hh\_type 6 3466.8 3496.8  
- lti 1 3505.7 3545.7  
- mh 1 3534.9 3574.9  
- ph 1 3624.5 3664.5  
  
Step: AIC=3450.78  
becomes\_eilts ~ age\_group + hascar + lti + mh + ph + hh\_type +   
 hiqual\_dv + eq\_net\_income  
  
 Df Deviance AIC  
<none> 3410.8 3450.8  
- eq\_net\_income 1 3419.4 3457.4  
- hascar 1 3423.0 3461.0  
- age\_group 3 3427.9 3461.9  
- hiqual\_dv 5 3452.3 3482.3  
- hh\_type 6 3467.6 3495.6  
- lti 1 3507.4 3545.4  
- mh 1 3535.4 3573.4  
- ph 1 3624.5 3662.5

Start: AIC=4548.77  
becomes\_eilts ~ 1  
  
 Df Deviance AIC  
+ lti 1 3943.8 3947.8  
+ ph 1 3955.5 3959.5  
+ mh 1 4288.6 4292.6  
+ hh\_type 6 4378.1 4392.1  
+ age\_group 3 4415.0 4423.0  
+ hiqual\_dv 5 4428.9 4440.9  
+ hascar 1 4477.3 4481.3  
<none> 4546.8 4548.8  
+ eth\_simplified 1 4544.9 4548.9  
+ eq\_net\_income 1 4546.4 4550.4  
+ sex 1 4546.7 4550.7  
  
Step: AIC=3947.85  
becomes\_eilts ~ lti  
  
 Df Deviance AIC  
+ ph 1 3727.5 3733.5  
+ mh 1 3845.1 3851.1  
+ hiqual\_dv 5 3858.6 3872.6  
+ hh\_type 6 3858.5 3874.5  
+ hascar 1 3897.4 3903.4  
+ age\_group 3 3894.9 3904.9  
<none> 3943.8 3947.8  
+ eq\_net\_income 1 3942.8 3948.8  
+ sex 1 3943.8 3949.8  
+ eth\_simplified 1 3943.8 3949.8  
  
Step: AIC=3733.46  
becomes\_eilts ~ lti + ph  
  
 Df Deviance AIC  
+ mh 1 3574.4 3582.4  
+ hh\_type 6 3633.1 3651.1  
+ hascar 1 3676.2 3684.2  
+ hiqual\_dv 5 3676.2 3692.2  
+ age\_group 3 3701.4 3713.4  
<none> 3727.5 3733.5  
+ eq\_net\_income 1 3726.8 3734.8  
+ eth\_simplified 1 3727.3 3735.3  
+ sex 1 3727.5 3735.5  
  
Step: AIC=3582.38  
becomes\_eilts ~ lti + ph + mh  
  
 Df Deviance AIC  
+ hh\_type 6 3495.4 3515.4  
+ hascar 1 3531.4 3541.4  
+ hiqual\_dv 5 3528.4 3546.4  
+ age\_group 3 3552.4 3566.4  
+ sex 1 3571.6 3581.6  
<none> 3574.4 3582.4  
+ eq\_net\_income 1 3573.6 3583.6  
+ eth\_simplified 1 3573.9 3583.9  
  
Step: AIC=3515.45  
becomes\_eilts ~ lti + ph + mh + hh\_type  
  
 Df Deviance AIC  
+ hiqual\_dv 5 3446.5 3476.5  
+ hascar 1 3477.3 3499.3  
+ age\_group 3 3479.8 3505.8  
+ eq\_net\_income 1 3490.1 3512.1  
<none> 3495.4 3515.4  
+ eth\_simplified 1 3494.7 3516.7  
+ sex 1 3495.2 3517.2  
  
Step: AIC=3476.49  
becomes\_eilts ~ lti + ph + mh + hh\_type + hiqual\_dv  
  
 Df Deviance AIC  
+ age\_group 3 3431.2 3467.2  
+ hascar 1 3436.7 3468.7  
+ eq\_net\_income 1 3438.2 3470.2  
<none> 3446.5 3476.5  
+ sex 1 3446.3 3478.3  
+ eth\_simplified 1 3446.5 3478.5  
  
Step: AIC=3467.16  
becomes\_eilts ~ lti + ph + mh + hh\_type + hiqual\_dv + age\_group  
  
 Df Deviance AIC  
+ hascar 1 3419.4 3457.4  
+ eq\_net\_income 1 3423.0 3461.0  
<none> 3431.2 3467.2  
+ sex 1 3431.0 3469.0  
+ eth\_simplified 1 3431.0 3469.0  
  
Step: AIC=3457.42  
becomes\_eilts ~ lti + ph + mh + hh\_type + hiqual\_dv + age\_group +   
 hascar  
  
 Df Deviance AIC  
+ eq\_net\_income 1 3410.8 3450.8  
<none> 3419.4 3457.4  
+ eth\_simplified 1 3419.0 3459.0  
+ sex 1 3419.3 3459.3  
  
Step: AIC=3450.78  
becomes\_eilts ~ lti + ph + mh + hh\_type + hiqual\_dv + age\_group +   
 hascar + eq\_net\_income  
  
 Df Deviance AIC  
<none> 3410.8 3450.8  
+ eth\_simplified 1 3410.2 3452.2  
+ sex 1 3410.6 3452.6

Start: AIC=4423.91  
becomes\_eilts ~ age\_group + sex + eth\_simplified  
  
 Df Deviance AIC  
+ lti 1 3894.5 3908.5  
+ ph 1 3907.8 3921.8  
+ mh 1 4148.9 4162.9  
+ hh\_type 6 4288.4 4312.4  
+ hiqual\_dv 5 4309.9 4331.9  
+ hascar 1 4318.7 4332.7  
- sex 1 4413.2 4423.2  
- eth\_simplified 1 4413.7 4423.7  
<none> 4411.9 4423.9  
+ eq\_net\_income 1 4411.6 4425.6  
- age\_group 3 4544.8 4550.8  
  
Step: AIC=3908.47  
becomes\_eilts ~ age\_group + sex + eth\_simplified + lti  
  
 Df Deviance AIC  
+ ph 1 3700.8 3716.8  
+ mh 1 3786.4 3802.4  
+ hiqual\_dv 5 3813.1 3837.1  
+ hh\_type 6 3818.6 3844.6  
+ hascar 1 3834.8 3850.8  
- eth\_simplified 1 3894.5 3906.5  
- sex 1 3894.8 3906.8  
<none> 3894.5 3908.5  
+ eq\_net\_income 1 3893.6 3909.6  
- age\_group 3 3943.8 3951.8  
- lti 1 4411.9 4423.9  
  
Step: AIC=3716.78  
becomes\_eilts ~ age\_group + sex + eth\_simplified + lti + ph  
  
 Df Deviance AIC  
+ mh 1 3549.7 3567.7  
+ hh\_type 6 3611.2 3639.2  
+ hascar 1 3643.2 3661.2  
+ hiqual\_dv 5 3648.8 3674.8  
- sex 1 3700.8 3714.8  
- eth\_simplified 1 3701.4 3715.4  
<none> 3700.8 3716.8  
+ eq\_net\_income 1 3700.2 3718.2  
- age\_group 3 3727.3 3737.3  
- ph 1 3894.5 3908.5  
- lti 1 3907.8 3921.8  
  
Step: AIC=3567.65  
becomes\_eilts ~ age\_group + sex + eth\_simplified + lti + ph +   
 mh  
  
 Df Deviance AIC  
+ hh\_type 6 3478.3 3508.3  
+ hascar 1 3499.5 3519.5  
+ hiqual\_dv 5 3505.2 3533.2  
- eth\_simplified 1 3550.4 3566.4  
- sex 1 3551.6 3567.6  
<none> 3549.7 3567.7  
+ eq\_net\_income 1 3548.6 3568.6  
- age\_group 3 3571.2 3583.2  
- lti 1 3657.6 3673.6  
- mh 1 3700.8 3716.8  
- ph 1 3786.4 3802.4  
  
Step: AIC=3508.33  
becomes\_eilts ~ age\_group + sex + eth\_simplified + lti + ph +   
 mh + hh\_type  
  
 Df Deviance AIC  
+ hiqual\_dv 5 3430.9 3470.9  
+ hascar 1 3456.1 3488.1  
+ eq\_net\_income 1 3472.9 3504.9  
- sex 1 3478.6 3506.6  
- eth\_simplified 1 3479.6 3507.6  
<none> 3478.3 3508.3  
- age\_group 3 3494.4 3518.4  
- hh\_type 6 3549.7 3567.7  
- lti 1 3573.3 3601.3  
- mh 1 3611.2 3639.2  
- ph 1 3728.0 3756.0  
  
Step: AIC=3470.87  
becomes\_eilts ~ age\_group + sex + eth\_simplified + lti + ph +   
 mh + hh\_type + hiqual\_dv  
  
 Df Deviance AIC  
+ hascar 1 3418.9 3460.9  
+ eq\_net\_income 1 3422.5 3464.5  
- eth\_simplified 1 3431.0 3469.0  
- sex 1 3431.0 3469.0  
<none> 3430.9 3470.9  
- age\_group 3 3446.3 3480.3  
- hiqual\_dv 5 3478.3 3508.3  
- hh\_type 6 3505.2 3533.2  
- lti 1 3527.5 3565.5  
- mh 1 3557.1 3595.1  
- ph 1 3647.8 3685.8  
  
Step: AIC=3460.87  
becomes\_eilts ~ age\_group + sex + eth\_simplified + lti + ph +   
 mh + hh\_type + hiqual\_dv + hascar  
  
 Df Deviance AIC  
+ eq\_net\_income 1 3410.0 3454.0  
- sex 1 3419.0 3459.0  
- eth\_simplified 1 3419.3 3459.3  
<none> 3418.9 3460.9  
- hascar 1 3430.9 3470.9  
- age\_group 3 3436.3 3472.3  
- hiqual\_dv 5 3456.1 3488.1  
- hh\_type 6 3469.9 3499.9  
- lti 1 3514.7 3554.7  
- mh 1 3543.6 3583.6  
- ph 1 3635.6 3675.6  
  
Step: AIC=3454.01  
becomes\_eilts ~ age\_group + sex + eth\_simplified + lti + ph +   
 mh + hh\_type + hiqual\_dv + hascar + eq\_net\_income  
  
 Df Deviance AIC  
- sex 1 3410.2 3452.2  
- eth\_simplified 1 3410.6 3452.6  
<none> 3410.0 3454.0  
- eq\_net\_income 1 3418.9 3460.9  
- hascar 1 3422.5 3464.5  
- age\_group 3 3427.3 3465.3  
- hiqual\_dv 5 3449.8 3483.8  
- hh\_type 6 3464.4 3496.4  
- lti 1 3505.3 3547.3  
- mh 1 3533.7 3575.7  
- ph 1 3624.5 3666.5  
  
Step: AIC=3452.24  
becomes\_eilts ~ age\_group + eth\_simplified + lti + ph + mh +   
 hh\_type + hiqual\_dv + hascar + eq\_net\_income  
  
 Df Deviance AIC  
- eth\_simplified 1 3410.8 3450.8  
<none> 3410.2 3452.2  
+ sex 1 3410.0 3454.0  
- eq\_net\_income 1 3419.0 3459.0  
- hascar 1 3422.8 3462.8  
- age\_group 3 3427.6 3463.6  
- hiqual\_dv 5 3450.0 3482.0  
- hh\_type 6 3466.8 3496.8  
- lti 1 3505.7 3545.7  
- mh 1 3534.9 3574.9  
- ph 1 3624.5 3664.5  
  
Step: AIC=3450.78  
becomes\_eilts ~ age\_group + lti + ph + mh + hh\_type + hiqual\_dv +   
 hascar + eq\_net\_income  
  
 Df Deviance AIC  
<none> 3410.8 3450.8  
+ eth\_simplified 1 3410.2 3452.2  
+ sex 1 3410.6 3452.6  
- eq\_net\_income 1 3419.4 3457.4  
- hascar 1 3423.0 3461.0  
- age\_group 3 3427.9 3461.9  
- hiqual\_dv 5 3452.3 3482.3  
- hh\_type 6 3467.6 3495.6  
- lti 1 3507.4 3545.4  
- mh 1 3535.4 3573.4  
- ph 1 3624.5 3662.5

Let’s look at the specifications arrived at by the three approaches

Call:  
glm(formula = becomes\_eilts ~ age\_group + hascar + lti + mh +   
 ph + hh\_type + hiqual\_dv + eq\_net\_income, family = binomial,   
 data = data\_tidied\_unemployed)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 8.877e-01 3.898e-01 2.277 0.022780  
age\_group25-44 7.574e-01 2.179e-01 3.477 0.000508  
age\_group45-54 8.538e-01 2.272e-01 3.759 0.000171  
age\_group55-64 8.260e-01 2.414e-01 3.422 0.000621  
hascar -3.590e-01 1.031e-01 -3.482 0.000498  
ltiyes 1.159e+00 1.234e-01 9.393 < 2e-16  
mh -4.102e-02 3.717e-03 -11.038 < 2e-16  
ph -5.792e-02 4.063e-03 -14.257 < 2e-16  
hh\_typeSmall Adult -5.687e-01 1.317e-01 -4.318 1.57e-05  
hh\_typeSingle Parent -5.648e-01 1.459e-01 -3.872 0.000108  
hh\_typeFamily with 1-2 Children -5.435e-01 1.499e-01 -3.627 0.000287  
hh\_typeFamily with 3 or more Children -9.157e-01 2.011e-01 -4.554 5.27e-06  
hh\_typeSingle Pensioner -1.012e+00 4.383e-01 -2.309 0.020971  
hh\_typePensioner Couple -1.575e+00 3.534e-01 -4.456 8.37e-06  
hiqual\_dvDegree -6.812e-01 2.080e-01 -3.275 0.001056  
hiqual\_dvGCSE etc -8.099e-02 1.412e-01 -0.573 0.566364  
hiqual\_dvNo qualification 4.286e-01 1.470e-01 2.915 0.003558  
hiqual\_dvOther higher degree -2.314e-01 1.893e-01 -1.222 0.221582  
hiqual\_dvOther qualification 1.889e-01 1.583e-01 1.194 0.232580  
eq\_net\_income 1.731e-04 4.833e-05 3.582 0.000341  
   
(Intercept) \*   
age\_group25-44 \*\*\*  
age\_group45-54 \*\*\*  
age\_group55-64 \*\*\*  
hascar \*\*\*  
ltiyes \*\*\*  
mh \*\*\*  
ph \*\*\*  
hh\_typeSmall Adult \*\*\*  
hh\_typeSingle Parent \*\*\*  
hh\_typeFamily with 1-2 Children \*\*\*  
hh\_typeFamily with 3 or more Children \*\*\*  
hh\_typeSingle Pensioner \*   
hh\_typePensioner Couple \*\*\*  
hiqual\_dvDegree \*\*   
hiqual\_dvGCSE etc   
hiqual\_dvNo qualification \*\*   
hiqual\_dvOther higher degree   
hiqual\_dvOther qualification   
eq\_net\_income \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 4546.8 on 7883 degrees of freedom  
Residual deviance: 3410.8 on 7864 degrees of freedom  
AIC: 3450.8  
  
Number of Fisher Scoring iterations: 6

20 variables included Now the forward approach

Call:  
glm(formula = becomes\_eilts ~ lti + ph + mh + hh\_type + hiqual\_dv +   
 age\_group + hascar + eq\_net\_income, family = binomial, data = data\_tidied\_unemployed)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 8.877e-01 3.898e-01 2.277 0.022780  
ltiyes 1.159e+00 1.234e-01 9.393 < 2e-16  
ph -5.792e-02 4.063e-03 -14.257 < 2e-16  
mh -4.102e-02 3.717e-03 -11.038 < 2e-16  
hh\_typeSmall Adult -5.687e-01 1.317e-01 -4.318 1.57e-05  
hh\_typeSingle Parent -5.648e-01 1.459e-01 -3.872 0.000108  
hh\_typeFamily with 1-2 Children -5.435e-01 1.499e-01 -3.627 0.000287  
hh\_typeFamily with 3 or more Children -9.157e-01 2.011e-01 -4.554 5.27e-06  
hh\_typeSingle Pensioner -1.012e+00 4.383e-01 -2.309 0.020971  
hh\_typePensioner Couple -1.575e+00 3.534e-01 -4.456 8.37e-06  
hiqual\_dvDegree -6.812e-01 2.080e-01 -3.275 0.001056  
hiqual\_dvGCSE etc -8.099e-02 1.412e-01 -0.573 0.566364  
hiqual\_dvNo qualification 4.286e-01 1.470e-01 2.915 0.003558  
hiqual\_dvOther higher degree -2.314e-01 1.893e-01 -1.222 0.221582  
hiqual\_dvOther qualification 1.889e-01 1.583e-01 1.194 0.232580  
age\_group25-44 7.574e-01 2.179e-01 3.477 0.000508  
age\_group45-54 8.538e-01 2.272e-01 3.759 0.000171  
age\_group55-64 8.260e-01 2.414e-01 3.422 0.000621  
hascar -3.590e-01 1.031e-01 -3.482 0.000498  
eq\_net\_income 1.731e-04 4.833e-05 3.582 0.000341  
   
(Intercept) \*   
ltiyes \*\*\*  
ph \*\*\*  
mh \*\*\*  
hh\_typeSmall Adult \*\*\*  
hh\_typeSingle Parent \*\*\*  
hh\_typeFamily with 1-2 Children \*\*\*  
hh\_typeFamily with 3 or more Children \*\*\*  
hh\_typeSingle Pensioner \*   
hh\_typePensioner Couple \*\*\*  
hiqual\_dvDegree \*\*   
hiqual\_dvGCSE etc   
hiqual\_dvNo qualification \*\*   
hiqual\_dvOther higher degree   
hiqual\_dvOther qualification   
age\_group25-44 \*\*\*  
age\_group45-54 \*\*\*  
age\_group55-64 \*\*\*  
hascar \*\*\*  
eq\_net\_income \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 4546.8 on 7883 degrees of freedom  
Residual deviance: 3410.8 on 7864 degrees of freedom  
AIC: 3450.8  
  
Number of Fisher Scoring iterations: 6

Now starting in the middle

Call:  
glm(formula = becomes\_eilts ~ age\_group + lti + ph + mh + hh\_type +   
 hiqual\_dv + hascar + eq\_net\_income, family = binomial, data = data\_tidied\_unemployed)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 8.877e-01 3.898e-01 2.277 0.022780  
age\_group25-44 7.574e-01 2.179e-01 3.477 0.000508  
age\_group45-54 8.538e-01 2.272e-01 3.759 0.000171  
age\_group55-64 8.260e-01 2.414e-01 3.422 0.000621  
ltiyes 1.159e+00 1.234e-01 9.393 < 2e-16  
ph -5.792e-02 4.063e-03 -14.257 < 2e-16  
mh -4.102e-02 3.717e-03 -11.038 < 2e-16  
hh\_typeSmall Adult -5.687e-01 1.317e-01 -4.318 1.57e-05  
hh\_typeSingle Parent -5.648e-01 1.459e-01 -3.872 0.000108  
hh\_typeFamily with 1-2 Children -5.435e-01 1.499e-01 -3.627 0.000287  
hh\_typeFamily with 3 or more Children -9.157e-01 2.011e-01 -4.554 5.27e-06  
hh\_typeSingle Pensioner -1.012e+00 4.383e-01 -2.309 0.020971  
hh\_typePensioner Couple -1.575e+00 3.534e-01 -4.456 8.37e-06  
hiqual\_dvDegree -6.812e-01 2.080e-01 -3.275 0.001056  
hiqual\_dvGCSE etc -8.099e-02 1.412e-01 -0.573 0.566364  
hiqual\_dvNo qualification 4.286e-01 1.470e-01 2.915 0.003558  
hiqual\_dvOther higher degree -2.314e-01 1.893e-01 -1.222 0.221582  
hiqual\_dvOther qualification 1.889e-01 1.583e-01 1.194 0.232580  
hascar -3.590e-01 1.031e-01 -3.482 0.000498  
eq\_net\_income 1.731e-04 4.833e-05 3.582 0.000341  
   
(Intercept) \*   
age\_group25-44 \*\*\*  
age\_group45-54 \*\*\*  
age\_group55-64 \*\*\*  
ltiyes \*\*\*  
ph \*\*\*  
mh \*\*\*  
hh\_typeSmall Adult \*\*\*  
hh\_typeSingle Parent \*\*\*  
hh\_typeFamily with 1-2 Children \*\*\*  
hh\_typeFamily with 3 or more Children \*\*\*  
hh\_typeSingle Pensioner \*   
hh\_typePensioner Couple \*\*\*  
hiqual\_dvDegree \*\*   
hiqual\_dvGCSE etc   
hiqual\_dvNo qualification \*\*   
hiqual\_dvOther higher degree   
hiqual\_dvOther qualification   
hascar \*\*\*  
eq\_net\_income \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 4546.8 on 7883 degrees of freedom  
Residual deviance: 3410.8 on 7864 degrees of freedom  
AIC: 3450.8  
  
Number of Fisher Scoring iterations: 6

Let’s get the Nagelkerke R^2 for each of these models

$N  
[1] 7884  
  
$R2  
[1] 0.3061868

$N  
[1] 7884  
  
$R2  
[1] 0.3061868

$N  
[1] 7884  
  
$R2  
[1] 0.3061868

Finally (finally?) let’s do the same for people who start off employed

Start: AIC=4892.08  
becomes\_eilts ~ age\_group + sex + hascar + eth\_simplified + lti +   
 mh + ph + hh\_type + hiqual\_dv + eq\_net\_income  
  
 Df Deviance AIC  
- age\_group 3 4851.7 4889.7  
- sex 1 4848.9 4890.9  
- eth\_simplified 1 4849.4 4891.4  
<none> 4848.1 4892.1  
- hh\_type 6 4862.4 4894.4  
- hascar 1 4856.5 4898.5  
- eq\_net\_income 1 4861.1 4903.1  
- lti 1 4877.6 4919.6  
- hiqual\_dv 5 4908.7 4942.7  
- mh 1 4970.9 5012.9  
- ph 1 5237.0 5279.0  
  
Step: AIC=4889.7  
becomes\_eilts ~ sex + hascar + eth\_simplified + lti + mh + ph +   
 hh\_type + hiqual\_dv + eq\_net\_income  
  
 Df Deviance AIC  
- sex 1 4852.3 4888.3  
- eth\_simplified 1 4853.0 4889.0  
<none> 4851.7 4889.7  
- hascar 1 4859.6 4895.6  
- hh\_type 6 4871.9 4897.9  
- eq\_net\_income 1 4864.6 4900.6  
- lti 1 4882.4 4918.4  
- hiqual\_dv 5 4917.3 4945.3  
- mh 1 4973.3 5009.3  
- ph 1 5246.2 5282.2  
  
Step: AIC=4888.34  
becomes\_eilts ~ hascar + eth\_simplified + lti + mh + ph + hh\_type +   
 hiqual\_dv + eq\_net\_income  
  
 Df Deviance AIC  
- eth\_simplified 1 4853.6 4887.6  
<none> 4852.3 4888.3  
- hascar 1 4860.5 4894.5  
- hh\_type 6 4872.4 4896.4  
- eq\_net\_income 1 4865.3 4899.3  
- lti 1 4883.0 4917.0  
- hiqual\_dv 5 4917.4 4943.4  
- mh 1 4977.0 5011.0  
- ph 1 5251.1 5285.1  
  
Step: AIC=4887.62  
becomes\_eilts ~ hascar + lti + mh + ph + hh\_type + hiqual\_dv +   
 eq\_net\_income  
  
 Df Deviance AIC  
<none> 4853.6 4887.6  
- hascar 1 4861.0 4893.0  
- hh\_type 6 4874.9 4896.9  
- eq\_net\_income 1 4866.3 4898.3  
- lti 1 4885.7 4917.7  
- hiqual\_dv 5 4920.3 4944.3  
- mh 1 4977.7 5009.7  
- ph 1 5251.2 5283.2

Start: AIC=5935.56  
becomes\_eilts ~ 1  
  
 Df Deviance AIC  
+ ph 1 5201.0 5205.0  
+ lti 1 5562.2 5566.2  
+ hiqual\_dv 5 5751.3 5763.3  
+ mh 1 5795.5 5799.5  
+ eq\_net\_income 1 5852.0 5856.0  
+ hh\_type 6 5844.5 5858.5  
+ age\_group 3 5861.3 5869.3  
+ hascar 1 5891.5 5895.5  
+ sex 1 5922.7 5926.7  
<none> 5933.6 5935.6  
+ eth\_simplified 1 5933.1 5937.1  
  
Step: AIC=5205.02  
becomes\_eilts ~ ph  
  
 Df Deviance AIC  
+ mh 1 5042.8 5048.8  
+ hiqual\_dv 5 5102.2 5116.2  
+ lti 1 5135.2 5141.2  
+ eq\_net\_income 1 5154.4 5160.4  
+ hh\_type 6 5157.1 5173.1  
+ hascar 1 5169.6 5175.6  
+ age\_group 3 5185.0 5195.0  
+ sex 1 5198.2 5204.2  
<none> 5201.0 5205.0  
+ eth\_simplified 1 5199.1 5205.1  
  
Step: AIC=5048.78  
becomes\_eilts ~ ph + mh  
  
 Df Deviance AIC  
+ hiqual\_dv 5 4939.1 4955.1  
+ lti 1 5004.9 5012.9  
+ eq\_net\_income 1 5008.4 5016.4  
+ hh\_type 6 5001.5 5019.5  
+ age\_group 3 5014.0 5026.0  
+ hascar 1 5018.4 5026.4  
+ eth\_simplified 1 5039.3 5047.3  
<none> 5042.8 5048.8  
+ sex 1 5042.6 5050.6  
  
Step: AIC=4955.12  
becomes\_eilts ~ ph + mh + hiqual\_dv  
  
 Df Deviance AIC  
+ lti 1 4902.0 4920.0  
+ hh\_type 6 4907.3 4935.3  
+ hascar 1 4922.0 4940.0  
+ eq\_net\_income 1 4927.1 4945.1  
+ age\_group 3 4927.8 4949.8  
<none> 4939.1 4955.1  
+ eth\_simplified 1 4937.4 4955.4  
+ sex 1 4937.9 4955.9  
  
Step: AIC=4920.01  
becomes\_eilts ~ ph + mh + hiqual\_dv + lti  
  
 Df Deviance AIC  
+ hascar 1 4884.4 4904.4  
+ hh\_type 6 4876.1 4906.1  
+ eq\_net\_income 1 4888.6 4908.6  
+ age\_group 3 4895.0 4919.0  
<none> 4902.0 4920.0  
+ sex 1 4900.9 4920.9  
+ eth\_simplified 1 4901.6 4921.6  
  
Step: AIC=4904.38  
becomes\_eilts ~ ph + mh + hiqual\_dv + lti + hascar  
  
 Df Deviance AIC  
+ eq\_net\_income 1 4874.9 4896.9  
+ hh\_type 6 4866.3 4898.3  
+ age\_group 3 4876.5 4902.5  
<none> 4884.4 4904.4  
+ eth\_simplified 1 4882.6 4904.6  
+ sex 1 4883.8 4905.8  
  
Step: AIC=4896.91  
becomes\_eilts ~ ph + mh + hiqual\_dv + lti + hascar + eq\_net\_income  
  
 Df Deviance AIC  
+ hh\_type 6 4853.6 4887.6  
+ age\_group 3 4865.0 4893.0  
+ eth\_simplified 1 4872.4 4896.4  
<none> 4874.9 4896.9  
+ sex 1 4874.3 4898.3  
  
Step: AIC=4887.62  
becomes\_eilts ~ ph + mh + hiqual\_dv + lti + hascar + eq\_net\_income +   
 hh\_type  
  
 Df Deviance AIC  
<none> 4853.6 4887.6  
+ eth\_simplified 1 4852.3 4888.3  
+ sex 1 4853.0 4889.0  
+ age\_group 3 4850.2 4890.2

Start: AIC=5860.42  
becomes\_eilts ~ age\_group + sex + eth\_simplified  
  
 Df Deviance AIC  
+ ph 1 5180.3 5194.3  
+ lti 1 5524.0 5538.0  
+ mh 1 5694.2 5708.2  
+ hiqual\_dv 5 5703.7 5725.7  
+ eq\_net\_income 1 5753.2 5767.2  
+ hascar 1 5800.9 5814.9  
+ hh\_type 6 5804.2 5828.2  
- eth\_simplified 1 5848.5 5858.5  
<none> 5848.4 5860.4  
- sex 1 5861.2 5871.2  
- age\_group 3 5922.3 5928.3  
  
Step: AIC=5194.28  
becomes\_eilts ~ age\_group + sex + eth\_simplified + ph  
  
 Df Deviance AIC  
+ mh 1 5011.7 5027.7  
+ hiqual\_dv 5 5091.5 5115.5  
+ lti 1 5122.5 5138.5  
+ eq\_net\_income 1 5126.8 5142.8  
+ hascar 1 5145.3 5161.3  
+ hh\_type 6 5147.8 5173.8  
- eth\_simplified 1 5181.2 5193.2  
<none> 5180.3 5194.3  
- sex 1 5184.1 5196.1  
- age\_group 3 5196.3 5204.3  
- ph 1 5848.4 5860.4  
  
Step: AIC=5027.72  
becomes\_eilts ~ age\_group + sex + eth\_simplified + ph + mh  
  
 Df Deviance AIC  
+ hiqual\_dv 5 4925.6 4951.6  
+ eq\_net\_income 1 4970.8 4988.8  
+ lti 1 4982.5 5000.5  
+ hascar 1 4983.6 5001.6  
+ hh\_type 6 4988.7 5016.7  
- sex 1 5012.1 5026.1  
- eth\_simplified 1 5013.6 5027.6  
<none> 5011.7 5027.7  
- age\_group 3 5039.1 5049.1  
- mh 1 5180.3 5194.3  
- ph 1 5694.2 5708.2  
  
Step: AIC=4951.61  
becomes\_eilts ~ age\_group + sex + eth\_simplified + ph + mh +   
 hiqual\_dv  
  
 Df Deviance AIC  
+ lti 1 4893.6 4921.6  
+ hascar 1 4905.8 4933.8  
+ eq\_net\_income 1 4909.4 4937.4  
+ hh\_type 6 4901.5 4939.5  
- eth\_simplified 1 4926.5 4950.5  
- sex 1 4926.9 4950.9  
<none> 4925.6 4951.6  
- age\_group 3 4936.2 4956.2  
- hiqual\_dv 5 5011.7 5027.7  
- mh 1 5091.5 5115.5  
- ph 1 5548.4 5572.4  
  
Step: AIC=4921.63  
becomes\_eilts ~ age\_group + sex + eth\_simplified + ph + mh +   
 hiqual\_dv + lti  
  
 Df Deviance AIC  
+ hascar 1 4874.5 4904.5  
+ eq\_net\_income 1 4877.5 4907.5  
+ hh\_type 6 4871.8 4911.8  
- eth\_simplified 1 4893.8 4919.8  
- sex 1 4894.8 4920.8  
<none> 4893.6 4921.6  
- age\_group 3 4900.5 4922.5  
- lti 1 4925.6 4951.6  
- hiqual\_dv 5 4982.5 5000.5  
- mh 1 5030.0 5056.0  
- ph 1 5295.2 5321.2  
  
Step: AIC=4904.53  
becomes\_eilts ~ age\_group + sex + eth\_simplified + ph + mh +   
 hiqual\_dv + lti + hascar  
  
 Df Deviance AIC  
+ eq\_net\_income 1 4862.4 4894.4  
+ hh\_type 6 4861.1 4903.1  
- sex 1 4875.2 4903.2  
- eth\_simplified 1 4875.9 4903.9  
<none> 4874.5 4904.5  
- age\_group 3 4882.0 4906.0  
- hascar 1 4893.6 4921.6  
- lti 1 4905.8 4933.8  
- hiqual\_dv 5 4955.2 4975.2  
- mh 1 5006.8 5034.8  
- ph 1 5275.8 5303.8  
  
Step: AIC=4894.41  
becomes\_eilts ~ age\_group + sex + eth\_simplified + ph + mh +   
 hiqual\_dv + lti + hascar + eq\_net\_income  
  
 Df Deviance AIC  
+ hh\_type 6 4848.1 4892.1  
- sex 1 4863.0 4893.0  
- eth\_simplified 1 4864.4 4894.4  
<none> 4862.4 4894.4  
- age\_group 3 4871.9 4897.9  
- eq\_net\_income 1 4874.5 4904.5  
- hascar 1 4877.5 4907.5  
- lti 1 4893.8 4923.8  
- hiqual\_dv 5 4922.6 4944.6  
- mh 1 4988.7 5018.7  
- ph 1 5252.5 5282.5  
  
Step: AIC=4892.08  
becomes\_eilts ~ age\_group + sex + eth\_simplified + ph + mh +   
 hiqual\_dv + lti + hascar + eq\_net\_income + hh\_type  
  
 Df Deviance AIC  
- age\_group 3 4851.7 4889.7  
- sex 1 4848.9 4890.9  
- eth\_simplified 1 4849.4 4891.4  
<none> 4848.1 4892.1  
- hh\_type 6 4862.4 4894.4  
- hascar 1 4856.5 4898.5  
- eq\_net\_income 1 4861.1 4903.1  
- lti 1 4877.6 4919.6  
- hiqual\_dv 5 4908.7 4942.7  
- mh 1 4970.9 5012.9  
- ph 1 5237.0 5279.0  
  
Step: AIC=4889.7  
becomes\_eilts ~ sex + eth\_simplified + ph + mh + hiqual\_dv +   
 lti + hascar + eq\_net\_income + hh\_type  
  
 Df Deviance AIC  
- sex 1 4852.3 4888.3  
- eth\_simplified 1 4853.0 4889.0  
<none> 4851.7 4889.7  
+ age\_group 3 4848.1 4892.1  
- hascar 1 4859.6 4895.6  
- hh\_type 6 4871.9 4897.9  
- eq\_net\_income 1 4864.6 4900.6  
- lti 1 4882.4 4918.4  
- hiqual\_dv 5 4917.3 4945.3  
- mh 1 4973.3 5009.3  
- ph 1 5246.2 5282.2  
  
Step: AIC=4888.34  
becomes\_eilts ~ eth\_simplified + ph + mh + hiqual\_dv + lti +   
 hascar + eq\_net\_income + hh\_type  
  
 Df Deviance AIC  
- eth\_simplified 1 4853.6 4887.6  
<none> 4852.3 4888.3  
+ sex 1 4851.7 4889.7  
+ age\_group 3 4848.9 4890.9  
- hascar 1 4860.5 4894.5  
- hh\_type 6 4872.4 4896.4  
- eq\_net\_income 1 4865.3 4899.3  
- lti 1 4883.0 4917.0  
- hiqual\_dv 5 4917.4 4943.4  
- mh 1 4977.0 5011.0  
- ph 1 5251.1 5285.1  
  
Step: AIC=4887.62  
becomes\_eilts ~ ph + mh + hiqual\_dv + lti + hascar + eq\_net\_income +   
 hh\_type  
  
 Df Deviance AIC  
<none> 4853.6 4887.6  
+ eth\_simplified 1 4852.3 4888.3  
+ sex 1 4853.0 4889.0  
+ age\_group 3 4850.2 4890.2  
- hascar 1 4861.0 4893.0  
- hh\_type 6 4874.9 4896.9  
- eq\_net\_income 1 4866.3 4898.3  
- lti 1 4885.7 4917.7  
- hiqual\_dv 5 4920.3 4944.3  
- mh 1 4977.7 5009.7  
- ph 1 5251.2 5283.2

Call:  
glm(formula = becomes\_eilts ~ hascar + lti + mh + ph + hh\_type +   
 hiqual\_dv + eq\_net\_income, family = binomial, data = data\_tidied\_employed)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 1.350e+00 3.454e-01 3.910 9.25e-05  
hascar -3.701e-01 1.330e-01 -2.782 0.005396  
ltiyes 6.635e-01 1.183e-01 5.610 2.02e-08  
mh -4.859e-02 4.200e-03 -11.569 < 2e-16  
ph -8.511e-02 4.126e-03 -20.624 < 2e-16  
hh\_typeSmall Adult -2.465e-01 1.420e-01 -1.737 0.082437  
hh\_typeSingle Parent -3.871e-01 2.037e-01 -1.900 0.057387  
hh\_typeFamily with 1-2 Children -5.707e-01 1.461e-01 -3.906 9.38e-05  
hh\_typeFamily with 3 or more Children -5.948e-01 2.267e-01 -2.624 0.008692  
hh\_typeSingle Pensioner -3.053e-01 3.400e-01 -0.898 0.369212  
hh\_typePensioner Couple -2.709e-03 1.896e-01 -0.014 0.988603  
hiqual\_dvDegree -2.513e-01 1.618e-01 -1.553 0.120462  
hiqual\_dvGCSE etc 3.410e-01 1.480e-01 2.304 0.021238  
hiqual\_dvNo qualification 1.120e+00 1.767e-01 6.342 2.26e-10  
hiqual\_dvOther higher degree 7.285e-02 1.748e-01 0.417 0.676852  
hiqual\_dvOther qualification 6.792e-01 1.723e-01 3.943 8.06e-05  
eq\_net\_income -2.136e-04 6.354e-05 -3.362 0.000774  
   
(Intercept) \*\*\*  
hascar \*\*   
ltiyes \*\*\*  
mh \*\*\*  
ph \*\*\*  
hh\_typeSmall Adult .   
hh\_typeSingle Parent .   
hh\_typeFamily with 1-2 Children \*\*\*  
hh\_typeFamily with 3 or more Children \*\*   
hh\_typeSingle Pensioner   
hh\_typePensioner Couple   
hiqual\_dvDegree   
hiqual\_dvGCSE etc \*   
hiqual\_dvNo qualification \*\*\*  
hiqual\_dvOther higher degree   
hiqual\_dvOther qualification \*\*\*  
eq\_net\_income \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 5933.6 on 118097 degrees of freedom  
Residual deviance: 4853.6 on 118081 degrees of freedom  
AIC: 4887.6  
  
Number of Fisher Scoring iterations: 9

Call:  
glm(formula = becomes\_eilts ~ ph + mh + hiqual\_dv + lti + hascar +   
 eq\_net\_income + hh\_type, family = binomial, data = data\_tidied\_employed)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 1.350e+00 3.454e-01 3.910 9.25e-05  
ph -8.511e-02 4.126e-03 -20.624 < 2e-16  
mh -4.859e-02 4.200e-03 -11.569 < 2e-16  
hiqual\_dvDegree -2.513e-01 1.618e-01 -1.553 0.120462  
hiqual\_dvGCSE etc 3.410e-01 1.480e-01 2.304 0.021238  
hiqual\_dvNo qualification 1.120e+00 1.767e-01 6.342 2.26e-10  
hiqual\_dvOther higher degree 7.285e-02 1.748e-01 0.417 0.676852  
hiqual\_dvOther qualification 6.792e-01 1.723e-01 3.943 8.06e-05  
ltiyes 6.635e-01 1.183e-01 5.610 2.02e-08  
hascar -3.701e-01 1.330e-01 -2.782 0.005396  
eq\_net\_income -2.136e-04 6.354e-05 -3.362 0.000774  
hh\_typeSmall Adult -2.465e-01 1.420e-01 -1.737 0.082437  
hh\_typeSingle Parent -3.871e-01 2.037e-01 -1.900 0.057387  
hh\_typeFamily with 1-2 Children -5.707e-01 1.461e-01 -3.906 9.38e-05  
hh\_typeFamily with 3 or more Children -5.948e-01 2.267e-01 -2.624 0.008692  
hh\_typeSingle Pensioner -3.053e-01 3.400e-01 -0.898 0.369212  
hh\_typePensioner Couple -2.709e-03 1.896e-01 -0.014 0.988603  
   
(Intercept) \*\*\*  
ph \*\*\*  
mh \*\*\*  
hiqual\_dvDegree   
hiqual\_dvGCSE etc \*   
hiqual\_dvNo qualification \*\*\*  
hiqual\_dvOther higher degree   
hiqual\_dvOther qualification \*\*\*  
ltiyes \*\*\*  
hascar \*\*   
eq\_net\_income \*\*\*  
hh\_typeSmall Adult .   
hh\_typeSingle Parent .   
hh\_typeFamily with 1-2 Children \*\*\*  
hh\_typeFamily with 3 or more Children \*\*   
hh\_typeSingle Pensioner   
hh\_typePensioner Couple   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 5933.6 on 118097 degrees of freedom  
Residual deviance: 4853.6 on 118081 degrees of freedom  
AIC: 4887.6  
  
Number of Fisher Scoring iterations: 9

Call:  
glm(formula = becomes\_eilts ~ ph + mh + hiqual\_dv + lti + hascar +   
 eq\_net\_income + hh\_type, family = binomial, data = data\_tidied\_employed)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 1.350e+00 3.454e-01 3.910 9.25e-05  
ph -8.511e-02 4.126e-03 -20.624 < 2e-16  
mh -4.859e-02 4.200e-03 -11.569 < 2e-16  
hiqual\_dvDegree -2.513e-01 1.618e-01 -1.553 0.120462  
hiqual\_dvGCSE etc 3.410e-01 1.480e-01 2.304 0.021238  
hiqual\_dvNo qualification 1.120e+00 1.767e-01 6.342 2.26e-10  
hiqual\_dvOther higher degree 7.285e-02 1.748e-01 0.417 0.676852  
hiqual\_dvOther qualification 6.792e-01 1.723e-01 3.943 8.06e-05  
ltiyes 6.635e-01 1.183e-01 5.610 2.02e-08  
hascar -3.701e-01 1.330e-01 -2.782 0.005396  
eq\_net\_income -2.136e-04 6.354e-05 -3.362 0.000774  
hh\_typeSmall Adult -2.465e-01 1.420e-01 -1.737 0.082437  
hh\_typeSingle Parent -3.871e-01 2.037e-01 -1.900 0.057387  
hh\_typeFamily with 1-2 Children -5.707e-01 1.461e-01 -3.906 9.38e-05  
hh\_typeFamily with 3 or more Children -5.948e-01 2.267e-01 -2.624 0.008692  
hh\_typeSingle Pensioner -3.053e-01 3.400e-01 -0.898 0.369212  
hh\_typePensioner Couple -2.709e-03 1.896e-01 -0.014 0.988603  
   
(Intercept) \*\*\*  
ph \*\*\*  
mh \*\*\*  
hiqual\_dvDegree   
hiqual\_dvGCSE etc \*   
hiqual\_dvNo qualification \*\*\*  
hiqual\_dvOther higher degree   
hiqual\_dvOther qualification \*\*\*  
ltiyes \*\*\*  
hascar \*\*   
eq\_net\_income \*\*\*  
hh\_typeSmall Adult .   
hh\_typeSingle Parent .   
hh\_typeFamily with 1-2 Children \*\*\*  
hh\_typeFamily with 3 or more Children \*\*   
hh\_typeSingle Pensioner   
hh\_typePensioner Couple   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 5933.6 on 118097 degrees of freedom  
Residual deviance: 4853.6 on 118081 degrees of freedom  
AIC: 4887.6  
  
Number of Fisher Scoring iterations: 9

Now Nagelkerke R^2 for each of these models

$N  
[1] 118098  
  
$R2  
[1] 0.1857658

$N  
[1] 118098  
  
$R2  
[1] 0.1857658

$N  
[1] 118098  
  
$R2  
[1] 0.1857658

Now, finally (finally (finally)) we can look at flows out of EILTS from those who start in this state

Start: AIC=6959.08  
leaves\_eilts ~ age\_group + sex + hascar + eth\_simplified + lti +   
 mh + ph + hh\_type + hiqual\_dv + eq\_net\_income  
  
 Df Deviance AIC  
- sex 1 6915.3 6957.3  
<none> 6915.1 6959.1  
- hiqual\_dv 5 6925.1 6959.1  
- eth\_simplified 1 6921.5 6963.5  
- eq\_net\_income 1 6921.8 6963.8  
- hascar 1 6933.4 6975.4  
- lti 1 6943.9 6985.9  
- age\_group 3 6979.6 7017.6  
- mh 1 6981.6 7023.6  
- hh\_type 6 6997.5 7029.5  
- ph 1 7097.8 7139.8  
  
Step: AIC=6957.35  
leaves\_eilts ~ age\_group + hascar + eth\_simplified + lti + mh +   
 ph + hh\_type + hiqual\_dv + eq\_net\_income  
  
 Df Deviance AIC  
- hiqual\_dv 5 6925.3 6957.3  
<none> 6915.3 6957.3  
- eth\_simplified 1 6921.7 6961.7  
- eq\_net\_income 1 6921.9 6961.9  
- hascar 1 6933.7 6973.7  
- lti 1 6944.2 6984.2  
- age\_group 3 6979.6 7015.6  
- mh 1 6981.9 7021.9  
- hh\_type 6 7002.1 7032.1  
- ph 1 7097.9 7137.9  
  
Step: AIC=6957.31  
leaves\_eilts ~ age\_group + hascar + eth\_simplified + lti + mh +   
 ph + hh\_type + eq\_net\_income  
  
 Df Deviance AIC  
<none> 6925.3 6957.3  
- eq\_net\_income 1 6930.8 6960.8  
- eth\_simplified 1 6932.5 6962.5  
- hascar 1 6948.3 6978.3  
- lti 1 6954.4 6984.4  
- age\_group 3 6988.2 7014.2  
- mh 1 6992.7 7022.7  
- hh\_type 6 7010.8 7030.8  
- ph 1 7109.5 7139.5

Start: AIC=7412.71  
leaves\_eilts ~ 1  
  
 Df Deviance AIC  
+ ph 1 7297.2 7301.2  
+ hh\_type 6 7293.9 7307.9  
+ age\_group 3 7315.7 7323.7  
+ lti 1 7333.6 7337.6  
+ mh 1 7375.6 7379.6  
+ hascar 1 7379.3 7383.3  
+ hiqual\_dv 5 7397.3 7409.3  
+ eq\_net\_income 1 7407.5 7411.5  
<none> 7410.7 7412.7  
+ eth\_simplified 1 7409.1 7413.1  
+ sex 1 7410.7 7414.7  
  
Step: AIC=7301.17  
leaves\_eilts ~ ph  
  
 Df Deviance AIC  
+ hh\_type 6 7139.3 7155.3  
+ age\_group 3 7196.1 7206.1  
+ mh 1 7204.0 7210.0  
+ hascar 1 7232.4 7238.4  
+ lti 1 7247.0 7253.0  
+ hiqual\_dv 5 7284.4 7298.4  
<none> 7297.2 7301.2  
+ eth\_simplified 1 7296.4 7302.4  
+ sex 1 7296.8 7302.8  
+ eq\_net\_income 1 7296.9 7302.9  
  
Step: AIC=7155.35  
leaves\_eilts ~ ph + hh\_type  
  
 Df Deviance AIC  
+ mh 1 7051.4 7069.4  
+ age\_group 3 7067.7 7089.7  
+ lti 1 7094.1 7112.1  
+ hascar 1 7118.7 7136.7  
+ eq\_net\_income 1 7135.0 7153.0  
+ hiqual\_dv 5 7127.1 7153.1  
+ eth\_simplified 1 7135.3 7153.3  
<none> 7139.3 7155.3  
+ sex 1 7138.1 7156.1  
  
Step: AIC=7069.44  
leaves\_eilts ~ ph + hh\_type + mh  
  
 Df Deviance AIC  
+ age\_group 3 6987.4 7011.4  
+ lti 1 7020.4 7040.4  
+ hascar 1 7031.7 7051.7  
+ eq\_net\_income 1 7046.6 7066.6  
+ eth\_simplified 1 7046.8 7066.8  
+ hiqual\_dv 5 7039.9 7067.9  
<none> 7051.4 7069.4  
+ sex 1 7051.4 7071.4  
  
Step: AIC=7011.37  
leaves\_eilts ~ ph + hh\_type + mh + age\_group  
  
 Df Deviance AIC  
+ lti 1 6956.7 6982.7  
+ hascar 1 6967.9 6993.9  
+ hiqual\_dv 5 6973.6 7007.6  
+ eth\_simplified 1 6982.4 7008.4  
+ eq\_net\_income 1 6983.2 7009.2  
<none> 6987.4 7011.4  
+ sex 1 6987.2 7013.2  
  
Step: AIC=6982.7  
leaves\_eilts ~ ph + hh\_type + mh + age\_group + lti  
  
 Df Deviance AIC  
+ hascar 1 6937.8 6965.8  
+ hiqual\_dv 5 6943.1 6979.1  
+ eth\_simplified 1 6951.7 6979.7  
+ eq\_net\_income 1 6953.3 6981.3  
<none> 6956.7 6982.7  
+ sex 1 6956.6 6984.6  
  
Step: AIC=6965.83  
leaves\_eilts ~ ph + hh\_type + mh + age\_group + lti + hascar  
  
 Df Deviance AIC  
+ eth\_simplified 1 6930.8 6960.8  
+ eq\_net\_income 1 6932.5 6962.5  
<none> 6937.8 6965.8  
+ hiqual\_dv 5 6928.3 6966.3  
+ sex 1 6937.8 6967.8  
  
Step: AIC=6960.76  
leaves\_eilts ~ ph + hh\_type + mh + age\_group + lti + hascar +   
 eth\_simplified  
  
 Df Deviance AIC  
+ eq\_net\_income 1 6925.3 6957.3  
<none> 6930.8 6960.8  
+ hiqual\_dv 5 6921.9 6961.9  
+ sex 1 6930.7 6962.7  
  
Step: AIC=6957.31  
leaves\_eilts ~ ph + hh\_type + mh + age\_group + lti + hascar +   
 eth\_simplified + eq\_net\_income  
  
 Df Deviance AIC  
<none> 6925.3 6957.3  
+ hiqual\_dv 5 6915.3 6957.3  
+ sex 1 6925.1 6959.1

Start: AIC=7325.77  
leaves\_eilts ~ age\_group + sex + eth\_simplified  
  
 Df Deviance AIC  
+ ph 1 7193.8 7207.8  
+ hh\_type 6 7219.5 7243.5  
+ lti 1 7238.4 7252.4  
+ hascar 1 7279.4 7293.4  
+ mh 1 7281.7 7295.7  
+ hiqual\_dv 5 7298.1 7320.1  
- sex 1 7313.9 7323.9  
+ eq\_net\_income 1 7310.4 7324.4  
- eth\_simplified 1 7315.6 7325.6  
<none> 7313.8 7325.8  
- age\_group 3 7409.1 7415.1  
  
Step: AIC=7207.8  
leaves\_eilts ~ age\_group + sex + eth\_simplified + ph  
  
 Df Deviance AIC  
+ hh\_type 6 7062.8 7088.8  
+ mh 1 7108.7 7124.7  
+ hascar 1 7124.9 7140.9  
+ lti 1 7143.8 7159.8  
+ hiqual\_dv 5 7176.2 7200.2  
- sex 1 7194.8 7206.8  
- eth\_simplified 1 7195.2 7207.2  
<none> 7193.8 7207.8  
+ eq\_net\_income 1 7193.4 7209.4  
- age\_group 3 7296.0 7304.0  
- ph 1 7313.8 7325.8  
  
Step: AIC=7088.76  
leaves\_eilts ~ age\_group + sex + eth\_simplified + ph + hh\_type  
  
 Df Deviance AIC  
+ mh 1 6982.3 7010.3  
+ lti 1 7018.4 7046.4  
+ hascar 1 7040.4 7068.4  
+ hiqual\_dv 5 7048.6 7084.6  
+ eq\_net\_income 1 7059.1 7087.1  
- sex 1 7063.2 7087.2  
<none> 7062.8 7088.8  
- eth\_simplified 1 7067.2 7091.2  
- age\_group 3 7134.1 7154.1  
- hh\_type 6 7193.8 7207.8  
- ph 1 7219.5 7243.5  
  
Step: AIC=7010.25  
leaves\_eilts ~ age\_group + sex + eth\_simplified + ph + hh\_type +   
 mh  
  
 Df Deviance AIC  
+ lti 1 6951.5 6981.5  
+ hascar 1 6960.9 6990.9  
+ hiqual\_dv 5 6969.2 7007.2  
+ eq\_net\_income 1 6977.9 7007.9  
- sex 1 6982.4 7008.4  
<none> 6982.3 7010.3  
- eth\_simplified 1 6987.2 7013.2  
- age\_group 3 7046.8 7068.8  
- mh 1 7062.8 7088.8  
- hh\_type 6 7108.7 7124.7  
- ph 1 7193.9 7219.9  
  
Step: AIC=6981.53  
leaves\_eilts ~ age\_group + sex + eth\_simplified + ph + hh\_type +   
 mh + lti  
  
 Df Deviance AIC  
+ hascar 1 6930.7 6962.7  
+ hiqual\_dv 5 6938.6 6978.6  
- sex 1 6951.7 6979.7  
+ eq\_net\_income 1 6948.0 6980.0  
<none> 6951.5 6981.5  
- eth\_simplified 1 6956.6 6984.6  
- lti 1 6982.3 7010.3  
- age\_group 3 7015.7 7039.7  
- mh 1 7018.4 7046.4  
- hh\_type 6 7073.5 7091.5  
- ph 1 7126.0 7154.0  
  
Step: AIC=6962.72  
leaves\_eilts ~ age\_group + sex + eth\_simplified + ph + hh\_type +   
 mh + lti + hascar  
  
 Df Deviance AIC  
+ eq\_net\_income 1 6925.1 6959.1  
- sex 1 6930.8 6960.8  
<none> 6930.7 6962.7  
+ hiqual\_dv 5 6921.8 6963.8  
- eth\_simplified 1 6937.8 6967.8  
- hascar 1 6951.5 6981.5  
- lti 1 6960.9 6990.9  
- age\_group 3 6994.3 7020.3  
- mh 1 6997.0 7027.0  
- hh\_type 6 7010.4 7030.4  
- ph 1 7118.6 7148.6  
  
Step: AIC=6959.14  
leaves\_eilts ~ age\_group + sex + eth\_simplified + ph + hh\_type +   
 mh + lti + hascar + eq\_net\_income  
  
 Df Deviance AIC  
- sex 1 6925.3 6957.3  
+ hiqual\_dv 5 6915.1 6959.1  
<none> 6925.1 6959.1  
- eq\_net\_income 1 6930.7 6962.7  
- eth\_simplified 1 6932.3 6964.3  
- hascar 1 6948.0 6980.0  
- lti 1 6954.2 6986.2  
- age\_group 3 6988.2 7016.2  
- mh 1 6992.3 7024.3  
- hh\_type 6 7006.6 7028.6  
- ph 1 7109.3 7141.3  
  
Step: AIC=6957.31  
leaves\_eilts ~ age\_group + eth\_simplified + ph + hh\_type + mh +   
 lti + hascar + eq\_net\_income  
  
 Df Deviance AIC  
<none> 6925.3 6957.3  
+ hiqual\_dv 5 6915.3 6957.3  
+ sex 1 6925.1 6959.1  
- eq\_net\_income 1 6930.8 6960.8  
- eth\_simplified 1 6932.5 6962.5  
- hascar 1 6948.3 6978.3  
- lti 1 6954.4 6984.4  
- age\_group 3 6988.2 7014.2  
- mh 1 6992.7 7022.7  
- hh\_type 6 7010.8 7030.8  
- ph 1 7109.5 7139.5

summaries

Call:  
glm(formula = leaves\_eilts ~ age\_group + hascar + eth\_simplified +   
 lti + mh + ph + hh\_type + eq\_net\_income, family = binomial,   
 data = data\_tidied\_eilts)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -2.0512063 0.3556240 -5.768 8.03e-09  
age\_group25-44 -0.2192683 0.2417255 -0.907 0.36436  
age\_group45-54 -0.5980716 0.2439143 -2.452 0.01421  
age\_group55-64 0.0126202 0.2455833 0.051 0.95902  
hascar 0.3332794 0.0696824 4.783 1.73e-06  
eth\_simplifiedWhite -0.2381612 0.0879782 -2.707 0.00679  
ltiyes -0.8911596 0.1652373 -5.393 6.92e-08  
mh 0.0199869 0.0024437 8.179 2.87e-16  
ph 0.0362892 0.0026867 13.507 < 2e-16  
hh\_typeSmall Adult 0.3671511 0.0885362 4.147 3.37e-05  
hh\_typeSingle Parent 0.5475556 0.1185595 4.618 3.87e-06  
hh\_typeFamily with 1-2 Children 0.5522897 0.1040624 5.307 1.11e-07  
hh\_typeFamily with 3 or more Children 0.4824734 0.1588577 3.037 0.00239  
hh\_typeSingle Pensioner 0.8279946 0.1510778 5.481 4.24e-08  
hh\_typePensioner Couple 0.7816120 0.1099629 7.108 1.18e-12  
eq\_net\_income -0.0001207 0.0000529 -2.282 0.02249  
   
(Intercept) \*\*\*  
age\_group25-44   
age\_group45-54 \*   
age\_group55-64   
hascar \*\*\*  
eth\_simplifiedWhite \*\*   
ltiyes \*\*\*  
mh \*\*\*  
ph \*\*\*  
hh\_typeSmall Adult \*\*\*  
hh\_typeSingle Parent \*\*\*  
hh\_typeFamily with 1-2 Children \*\*\*  
hh\_typeFamily with 3 or more Children \*\*   
hh\_typeSingle Pensioner \*\*\*  
hh\_typePensioner Couple \*\*\*  
eq\_net\_income \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 7410.7 on 6548 degrees of freedom  
Residual deviance: 6925.3 on 6533 degrees of freedom  
AIC: 6957.3  
  
Number of Fisher Scoring iterations: 4

Next model summary

Call:  
glm(formula = leaves\_eilts ~ ph + hh\_type + mh + age\_group +   
 lti + hascar + eth\_simplified + eq\_net\_income, family = binomial,   
 data = data\_tidied\_eilts)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -2.0512063 0.3556240 -5.768 8.03e-09  
ph 0.0362892 0.0026867 13.507 < 2e-16  
hh\_typeSmall Adult 0.3671511 0.0885362 4.147 3.37e-05  
hh\_typeSingle Parent 0.5475556 0.1185595 4.618 3.87e-06  
hh\_typeFamily with 1-2 Children 0.5522897 0.1040624 5.307 1.11e-07  
hh\_typeFamily with 3 or more Children 0.4824734 0.1588577 3.037 0.00239  
hh\_typeSingle Pensioner 0.8279946 0.1510778 5.481 4.24e-08  
hh\_typePensioner Couple 0.7816120 0.1099629 7.108 1.18e-12  
mh 0.0199869 0.0024437 8.179 2.87e-16  
age\_group25-44 -0.2192683 0.2417255 -0.907 0.36436  
age\_group45-54 -0.5980716 0.2439143 -2.452 0.01421  
age\_group55-64 0.0126202 0.2455833 0.051 0.95902  
ltiyes -0.8911596 0.1652373 -5.393 6.92e-08  
hascar 0.3332794 0.0696824 4.783 1.73e-06  
eth\_simplifiedWhite -0.2381612 0.0879782 -2.707 0.00679  
eq\_net\_income -0.0001207 0.0000529 -2.282 0.02249  
   
(Intercept) \*\*\*  
ph \*\*\*  
hh\_typeSmall Adult \*\*\*  
hh\_typeSingle Parent \*\*\*  
hh\_typeFamily with 1-2 Children \*\*\*  
hh\_typeFamily with 3 or more Children \*\*   
hh\_typeSingle Pensioner \*\*\*  
hh\_typePensioner Couple \*\*\*  
mh \*\*\*  
age\_group25-44   
age\_group45-54 \*   
age\_group55-64   
ltiyes \*\*\*  
hascar \*\*\*  
eth\_simplifiedWhite \*\*   
eq\_net\_income \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 7410.7 on 6548 degrees of freedom  
Residual deviance: 6925.3 on 6533 degrees of freedom  
AIC: 6957.3  
  
Number of Fisher Scoring iterations: 4

Finally, both ways

Call:  
glm(formula = leaves\_eilts ~ age\_group + eth\_simplified + ph +   
 hh\_type + mh + lti + hascar + eq\_net\_income, family = binomial,   
 data = data\_tidied\_eilts)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -2.0512063 0.3556240 -5.768 8.03e-09  
age\_group25-44 -0.2192683 0.2417255 -0.907 0.36436  
age\_group45-54 -0.5980716 0.2439143 -2.452 0.01421  
age\_group55-64 0.0126202 0.2455833 0.051 0.95902  
eth\_simplifiedWhite -0.2381612 0.0879782 -2.707 0.00679  
ph 0.0362892 0.0026867 13.507 < 2e-16  
hh\_typeSmall Adult 0.3671511 0.0885362 4.147 3.37e-05  
hh\_typeSingle Parent 0.5475556 0.1185595 4.618 3.87e-06  
hh\_typeFamily with 1-2 Children 0.5522897 0.1040624 5.307 1.11e-07  
hh\_typeFamily with 3 or more Children 0.4824734 0.1588577 3.037 0.00239  
hh\_typeSingle Pensioner 0.8279946 0.1510778 5.481 4.24e-08  
hh\_typePensioner Couple 0.7816120 0.1099629 7.108 1.18e-12  
mh 0.0199869 0.0024437 8.179 2.87e-16  
ltiyes -0.8911596 0.1652373 -5.393 6.92e-08  
hascar 0.3332794 0.0696824 4.783 1.73e-06  
eq\_net\_income -0.0001207 0.0000529 -2.282 0.02249  
   
(Intercept) \*\*\*  
age\_group25-44   
age\_group45-54 \*   
age\_group55-64   
eth\_simplifiedWhite \*\*   
ph \*\*\*  
hh\_typeSmall Adult \*\*\*  
hh\_typeSingle Parent \*\*\*  
hh\_typeFamily with 1-2 Children \*\*\*  
hh\_typeFamily with 3 or more Children \*\*   
hh\_typeSingle Pensioner \*\*\*  
hh\_typePensioner Couple \*\*\*  
mh \*\*\*  
ltiyes \*\*\*  
hascar \*\*\*  
eq\_net\_income \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 7410.7 on 6548 degrees of freedom  
Residual deviance: 6925.3 on 6533 degrees of freedom  
AIC: 6957.3  
  
Number of Fisher Scoring iterations: 4

Now Nagelkerke R^2 for each of these models

$N  
[1] 6549  
  
$R2  
[1] 0.1054462

$N  
[1] 6549  
  
$R2  
[1] 0.1054462

$N  
[1] 6549  
  
$R2  
[1] 0.1054462

CODE BEGINS HERE

title: "Block-based models predicting flows into EILTS"

author:

- "Jon Minton"

- "Martin Taulbut"

format:

html:

warning: false

code-fold: true

message: false

code-summary: "Show R Code"

docx:

warning: false

echo: false

message: false

editor: visual

prefer-html: true

---

## Aim

This document will develop analyses which predict flows into EILTS using a block-based approach. This will involve developing a series of logistic regression models which predict flows into EILTS using a range of individual and household level attributes.

## Preparation

```{r}

devtools::load\_all(here::here("R"))

base\_dir\_location <- "big\_data/UKDA-6614-stata/stata/stata13\_se/ukhls"

library(tidyverse)

library(nnet)

library(knitr)

library(kableExtra)

# Individual level attributes

varnames <- c(

"jbstat", "dvage", "sex", # econ status; age; sex

"sf12mcs\_dv", "sf12pcs\_dv", # SF12 MH, SF12 PH

"health" # LLTI as a binary variable

)

extract\_what <- c(

"labels", "values", "labels",

"values", "values",

"labels"

)

demohealth\_data <- get\_ind\_level\_vars\_for\_selected\_waves(

varnames = varnames, vartypes = extract\_what

)

mh\_data <-

get\_ind\_level\_vars\_for\_selected\_waves(

varnames = c("jbstat", "sf12mcs\_dv"),

vartypes = c("labels", "values")

)

eth\_data <-

get\_ind\_level\_vars\_for\_selected\_waves(

varnames = c("jbstat", "ethn\_dv"),

vartypes = c("labels", "labels")

)

hiqual\_data <-

get\_ind\_level\_vars\_for\_selected\_waves(

varnames = c("jbstat", "hiqual\_dv"),

vartypes = c("labels", "labels")

)

gor\_data <-

get\_ind\_level\_vars\_for\_selected\_waves(

varnames = c("jbstat", "gor\_dv"),

vartypes = c("labels", "labels")

)

ind\_joined <-

left\_join(

demohealth\_data |> select(-sf12mcs\_dv),

eth\_data |> select(

pidp, wave, ethn\_dv

)

) |>

left\_join(

mh\_data |> select(

pidp, wave, sf12mcs\_dv

)

) |>

left\_join(

hiqual\_data |>

select(pidp, wave, hiqual\_dv)

) |>

left\_join(

gor\_data |>

select(pidp, wave, gor\_dv)

)

# Household level attributes

# cars: ncars

# equivalised hh income

# hh type

vars\_types\_hh <- tribble(

~var\_name, ~var\_type,

"fihhmnnet1\_dv", "values",

"ieqmoecd\_dv", "values",

"hhtype\_dv", "labels"

)

dta\_joined\_pt1 <-

add\_hh\_variables(

ind\_joined,

varnames = vars\_types\_hh$var\_name,

vartypes = vars\_types\_hh$var\_type

)

dta\_joined\_pt2 <-

add\_hh\_variables(

ind\_joined,

varnames = "ncars",

vartypes = "values"

)

dta\_joined <-

left\_join(

dta\_joined\_pt1,

dta\_joined\_pt2 |> select(pidp, wave, hidp, ncars)

)

rm(dta\_joined\_pt1, dta\_joined\_pt2, ind\_joined, mh\_data, eth\_data, demohealth\_data, hiqual\_data, gor\_data)

```

## Modelling

We will start with the simplest possible model specification, then add blocks of covariates

We are predicting whether next\_status is EILTS or not.

We will start with a manual approach to deciding on which blocks of variables, and variables within, to include

```{r}

data\_tidied <-

dta\_joined |>

mutate(

eth\_simplified = case\_when(

ethn\_dv %in% c("british/english/scottish/welsh/northern irish", "any other white background", "gypsy or irish traveller") ~ 'White',

is.na(ethn\_dv) | ethn\_dv == 'missing' ~ NA\_character\_,

TRUE ~ 'non-White'

)

) |>

mutate(

age = ifelse(dvage < 0, NA, dvage)

) |>

filter(between(age, 16, 64)) |>

filter(hiqual\_dv != 'missing') |>

mutate(

mh = ifelse(sf12mcs\_dv < 0, NA, sf12mcs\_dv),

ph = ifelse(sf12pcs\_dv < 0, NA, sf12pcs\_dv)

) |>

mutate(

ncars = ifelse(ncars < 0, NA, ncars),

hascar = ifelse(ncars > 0, 1, 0)

) |>

mutate(ieqmoecd\_dv = ifelse(ieqmoecd\_dv < 0, NA, ieqmoecd\_dv)) |>

mutate(

eq\_net\_income = fihhmnnet1\_dv / ieqmoecd\_dv

) |>

mutate(

ncars = ifelse(ncars < 0, NA, ncars)

) |>

mutate(

lti = case\_when(

health == '1' ~ 'yes',

health == '2' ~ 'no',

TRUE ~ NA\_character\_

)

) |>

left\_join(simplified\_household\_lookup, by = c('hhtype\_dv' = 'original')) |>

filter(!is.na(this\_status)) |>

filter(wave %in% letters[1:10]) |> # waves a to j

mutate(

becomes\_eilts = ifelse(next\_status == "Inactive long term sick", 1, 0)

) |>

mutate(

age\_group = case\_when(

between(age, 16, 24) ~ "16-24",

between(age, 25, 44) ~ "25-44",

between(age, 45, 54) ~ "45-54",

between(age, 55, 64) ~ "55-64"

)

) |>

filter(sex != 'missing') |>

mutate(

hh\_type = factor(recoded, levels = c("Single Adult", "Small Adult", "Single Parent", "Family with 1-2 Children", "Family with 3 or more Children", "Single Pensioner", "Pensioner Couple")

)

) |>

mutate(hiqual\_dv = case\_when(

hiqual\_dv %in% c("A level etc", "A-level etc") ~ "A level etc",

hiqual\_dv %in% c("No qual", "No qualification") ~ "No qualification",

hiqual\_dv %in% c("Other higher", "Other higher degree") ~ "Other higher degree",

hiqual\_dv %in% c("Other qual", "Other qualification") ~ "Other qualification",

hiqual\_dv %in% c("inapplicable", "missing") ~ NA\_character\_,

TRUE ~ hiqual\_dv

)

)

```

```{r}

mod\_null <- glm(becomes\_eilts ~ 1, data = data\_tidied, family = binomial)

summary(mod\_null)

```

Now to add the first block of variables: history

```{r}

mod\_history <- glm(becomes\_eilts ~ this\_status, data = data\_tidied, family = binomial)

summary(mod\_history)

```

Now to add the second block of variables: demographics

```{r}

mod\_history\_demographics <- glm(becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified, data = data\_tidied, family = binomial)

summary(mod\_history\_demographics)

```

We can now start comparing the AICs

```{r}

AIC(mod\_null, mod\_history, mod\_history\_demographics)

```

Now hh income, both linear and logged

```{r}

mod\_history\_demographics\_hhincome <- glm(

becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified + eq\_net\_income,

data = data\_tidied,

family = binomial

)

summary(mod\_history\_demographics\_hhincome)

```

Now logged

```{r}

mod\_history\_demographics\_loghhincome <- glm(

becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified + log(eq\_net\_income+ 0.5),

data = data\_tidied,

family = binomial

)

summary(mod\_history\_demographics\_loghhincome)

```

If we were to remove event history, the effect of income may be stat sig

```{r}

mod\_demographics\_loghhincome <- glm(

becomes\_eilts ~ age\_group + sex + eth\_simplified + log(eq\_net\_income + 0.5),

data = data\_tidied,

family = binomial

)

summary(mod\_demographics\_loghhincome)

```

Now car access

```{r}

mod\_history\_demographics\_hascar <- glm(

becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified + hascar,

data = data\_tidied,

family = binomial

)

summary(mod\_history\_demographics\_hascar)

```

Now to add the next block of variables: health

```{r}

mod\_history\_demographics\_car\_lti <- glm(

becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified + hascar + lti,

data = data\_tidied,

family = binomial

)

summary(mod\_history\_demographics\_car\_lti)

```

Now health as continuous variables

```{r}

mod\_history\_demographics\_car\_sf12 <- glm(

becomes\_eilts ~ this\_status + age\_group + sex + hascar + eth\_simplified + mh + ph,

data = data\_tidied,

family = binomial

)

summary(mod\_history\_demographics\_car\_sf12)

```

Finally let's look at both health and lti

```{r}

mod\_history\_demographics\_car\_health <- glm(

becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified + hascar + mh + ph + lti,

data = data\_tidied,

family = binomial

)

summary(mod\_history\_demographics\_car\_health)

```

But is the AIC of including both types of health better than just sf12 or lti?

```{r}

AIC(mod\_history\_demographics\_car\_sf12, mod\_history\_demographics\_car\_lti, mod\_history\_demographics\_car\_health)

```

The number of observations aren't exactly the same, so we can't directly compare AICs. However, the AIC of the model with both types of health is lower than the AIC of the model with just one type of health.

However it appears that the sf12 derived variables are more useful than the lti/no lti binary variable

So let's use mh and ph, before moving onto the next block of variables: household

```{r}

mod\_history\_demographics\_health\_car\_hhchildren <- glm(

becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified + lti + mh + ph + has\_children,

data = data\_tidied,

family = binomial

)

summary(mod\_history\_demographics\_health\_car\_hhchildren)

```

Now hh category

```{r}

mod\_history\_demographics\_health\_car\_hhtype <- glm(

becomes\_eilts ~ this\_status + age\_group + sex + hascar + eth\_simplified + lti + mh + ph + hh\_type,

data = data\_tidied,

family = binomial

)

summary(mod\_history\_demographics\_health\_car\_hhtype)

```

Now qualifications, which we think is the last block

```{r}

mod\_history\_demographics\_health\_car\_hhtype\_qual <- glm(

becomes\_eilts ~ this\_status + age\_group + sex + hascar + eth\_simplified + lti + mh + ph + hh\_type + hiqual\_dv,

data = data\_tidied,

family = binomial

)

summary(mod\_history\_demographics\_health\_car\_hhtype\_qual)

broom::tidy(mod\_history\_demographics\_health\_car\_hhtype\_qual) |> write.csv("final\_model\_results\_preds\_into\_eilts.csv")

```

We now have a series of models, organised into blocks, which build up in complexity incrementally. Each appears to increase the proportion explained. Let's use Nagelkerke's R^2 to compare the models

```{r}

fmsb::NagelkerkeR2(mod\_null)

fmsb::NagelkerkeR2(mod\_history)

fmsb::NagelkerkeR2(mod\_history\_demographics)

fmsb::NagelkerkeR2(mod\_history\_demographics\_hascar)

fmsb::NagelkerkeR2(mod\_history\_demographics\_car\_health)

fmsb::NagelkerkeR2(mod\_history\_demographics\_health\_car\_hhchildren) # slightly worse than previous

fmsb::NagelkerkeR2(mod\_history\_demographics\_health\_car\_hhtype)

fmsb::NagelkerkeR2(mod\_history\_demographics\_health\_car\_hhtype\_qual)

```

Now to make this a table

```{r}

label\_and\_return\_nr2 <- function(label, model){

nr2 <- fmsb::NagelkerkeR2(model)

tibble(label = label, nr2 = nr2$R2)

}

nr2s <- bind\_rows(

label\_and\_return\_nr2("null", mod\_null),

label\_and\_return\_nr2("history", mod\_history),

label\_and\_return\_nr2("history\_demographics", mod\_history\_demographics),

label\_and\_return\_nr2("history\_demographics\_hascar", mod\_history\_demographics\_hascar),

label\_and\_return\_nr2("history\_demographics\_car\_health", mod\_history\_demographics\_car\_health),

label\_and\_return\_nr2("history\_demographics\_health\_car\_hhchildren", mod\_history\_demographics\_health\_car\_hhchildren),

label\_and\_return\_nr2("history\_demographics\_health\_car\_hhtype", mod\_history\_demographics\_health\_car\_hhtype),

label\_and\_return\_nr2("history\_demographics\_health\_car\_hhtype\_qual", mod\_history\_demographics\_health\_car\_hhtype\_qual)

)

nr2s |> write.csv("nagelkerke\_r2s.csv")

nr2s |> kable() |> kable\_styling()

```

Let's return to the demographics block and see if we can do better (using the spec previously arrived at )

```{r}

mod\_history\_demographics\_better <- glm(becomes\_eilts ~ this\_status \* sex + splines::bs(age, 5) + sex + eth\_simplified, data = data\_tidied, family = binomial)

summary(mod\_history\_demographics\_better)

fmsb::NagelkerkeR2(mod\_history\_demographics\_better)

```

Let's now look at the stepwise AIC approach to see if similar variables are selected

```{r}

complete\_data <- data\_tidied %>% filter(complete.cases(.))

new\_mod\_null <- glm(becomes\_eilts ~ 1, data = complete\_data, family = binomial)

new\_mod\_history\_demographics <- glm(becomes\_eilts ~ this\_status + age\_group + sex + eth\_simplified, data = complete\_data, family = binomial)

new\_mod\_history\_demographics\_health\_car\_hhtype\_qual <- glm(

becomes\_eilts ~ this\_status + age\_group + sex + hascar + eth\_simplified + lti + mh + ph + hh\_type + hiqual\_dv,

data = complete\_data,

family = binomial

)

mdl\_stepAIC <- step(new\_mod\_history\_demographics, direction = "both", scope = list(lower = new\_mod\_null, upper = new\_mod\_history\_demographics\_health\_car\_hhtype\_qual))

```

Let's see the final specification arrived at

```{r}

summary(mdl\_stepAIC)

```

Let's see what happens if we ask the algorithm to prune our final model specification

```{r}

mdl\_stepAIC\_pruned <- step(new\_mod\_history\_demographics\_health\_car\_hhtype\_qual, direction = "backward", scope = list(lower = new\_mod\_null, upper = new\_mod\_history\_demographics\_health\_car\_hhtype\_qual))

summary(mdl\_stepAIC\_pruned)

```

The pruned model is the same as the model we derived manually, so all variables appear essential

Finally, we can start with the simplest and move forwards

```{r}

mdl\_stepAIC\_grown <- step(new\_mod\_null, direction = "forward", scope = list(lower = new\_mod\_null, upper = new\_mod\_history\_demographics\_health\_car\_hhtype\_qual))

summary(mdl\_stepAIC\_grown)

```

Once again we end up with the same model as we derived manually, so all variables appear essential.

Finally, for now, we'll look at just those who start off in the unemployed category. This means the history variable is no longer needed as everyone's history is now the same

```{r}

data\_tidied\_unemployed <- data\_tidied |> filter(this\_status == "Unemployed") %>% filter(complete.cases(.))

mod\_unemployed\_null <- glm(becomes\_eilts ~ 1, data = data\_tidied\_unemployed, family = binomial)

mod\_unemployed\_demographics <- glm(becomes\_eilts ~ age\_group + sex + eth\_simplified, data = data\_tidied\_unemployed, family = binomial)

mod\_unemployed\_full <- glm(becomes\_eilts ~ age\_group + sex + hascar + eth\_simplified + lti + mh + ph + hh\_type + hiqual\_dv + eq\_net\_income, data = data\_tidied\_unemployed, family = binomial) # this time with income

# first step backwards with the full model,

mdl\_stepAIC\_unemployed\_pruned <- step(mod\_unemployed\_full, direction = "backward", scope = list(lower = mod\_unemployed\_null, upper = mod\_unemployed\_full))

# then step forwards from the null model

mdl\_stepAIC\_unemployed\_grown <- step(mod\_unemployed\_null, direction = "forward", scope = list(lower = mod\_unemployed\_null, upper = mod\_unemployed\_full))

# then allow stepping in either direction with the demographics model to start

mdl\_stepAIC\_unemployed\_both <- step(mod\_unemployed\_demographics, direction = "both", scope = list(lower = mod\_unemployed\_null, upper = mod\_unemployed\_full))

```

Let's look at the specifications arrived at by the three approaches

```{r}

summary(mdl\_stepAIC\_unemployed\_pruned)

```

20 variables included

Now the forward approach

```{r}

summary(mdl\_stepAIC\_unemployed\_grown)

```

Now starting in the middle

```{r}

summary(mdl\_stepAIC\_unemployed\_both)

```

Let's get the Nagelkerke R^2 for each of these models

```{r}

fmsb::NagelkerkeR2(mdl\_stepAIC\_unemployed\_pruned)

fmsb::NagelkerkeR2(mdl\_stepAIC\_unemployed\_grown)

fmsb::NagelkerkeR2(mdl\_stepAIC\_unemployed\_both)

```

Finally (finally?) let's do the same for people who start off employed

```{r}

data\_tidied\_employed <- data\_tidied |> filter(this\_status == "Employed") %>% filter(complete.cases(.))

mod\_employed\_null <- glm(becomes\_eilts ~ 1, data = data\_tidied\_employed, family = binomial)

mod\_employed\_demographics <- glm(becomes\_eilts ~ age\_group + sex + eth\_simplified, data = data\_tidied\_employed, family = binomial)

mod\_employed\_full <- glm(becomes\_eilts ~ age\_group + sex + hascar + eth\_simplified + lti + mh + ph + hh\_type + hiqual\_dv + eq\_net\_income, data = data\_tidied\_employed, family = binomial) # this time with income

# first step backwards with the full model,

mdl\_stepAIC\_employed\_pruned <- step(mod\_employed\_full, direction = "backward", scope = list(lower = mod\_employed\_null, upper = mod\_employed\_full))

# then step forwards from the null model

mdl\_stepAIC\_employed\_grown <- step(mod\_employed\_null, direction = "forward", scope = list(lower = mod\_employed\_null, upper = mod\_employed\_full))

# then allow stepping in either direction with the demographics model to start

mdl\_stepAIC\_employed\_both <- step(mod\_employed\_demographics, direction = "both", scope = list(lower = mod\_employed\_null, upper = mod\_employed\_full))

```

```{r}

summary(mdl\_stepAIC\_employed\_pruned)

```

```{r}

summary(mdl\_stepAIC\_employed\_grown)

```

```{r}

summary(mdl\_stepAIC\_employed\_both)

```

Now Nagelkerke R^2 for each of these models

```{r}

fmsb::NagelkerkeR2(mdl\_stepAIC\_employed\_pruned)

fmsb::NagelkerkeR2(mdl\_stepAIC\_employed\_grown)

fmsb::NagelkerkeR2(mdl\_stepAIC\_employed\_both)

```

Now, finally (finally (finally)) we can look at flows out of EILTS from those who start in this state

```{r}

data\_tidied\_eilts <- data\_tidied |> filter(this\_status == "Inactive long term sick") %>% filter(complete.cases(.)) |>

mutate(leaves\_eilts = as.numeric(!becomes\_eilts))

mod\_eilts\_null <- glm(leaves\_eilts ~ 1, data = data\_tidied\_eilts, family = binomial)

mod\_eilts\_demographics <- glm(leaves\_eilts ~ age\_group + sex + eth\_simplified, data = data\_tidied\_eilts, family = binomial)

mod\_eilts\_full <- glm(leaves\_eilts ~ age\_group + sex + hascar + eth\_simplified + lti + mh + ph + hh\_type + hiqual\_dv + eq\_net\_income, data = data\_tidied\_eilts, family = binomial) # this time with income

# first step backwards with the full model,

mdl\_stepAIC\_eilts\_pruned <- step(mod\_eilts\_full, direction = "backward", scope = list(lower = mod\_eilts\_null, upper = mod\_eilts\_full))

# then step forwards from the null model

mdl\_stepAIC\_eilts\_grown <- step(mod\_eilts\_null, direction = "forward", scope = list(lower = mod\_eilts\_null, upper = mod\_eilts\_full))

# then allow stepping in either direction with the demographics model to start

mdl\_stepAIC\_eilts\_both <- step(mod\_eilts\_demographics, direction = "both", scope = list(lower = mod\_eilts\_null, upper = mod\_eilts\_full))

```

summaries

```{r}

summary(mdl\_stepAIC\_eilts\_pruned)

```

Next model summary

```{r}

summary(mdl\_stepAIC\_eilts\_grown)

```

Finally, both ways

```{r}

summary(mdl\_stepAIC\_eilts\_both)

```

Now Nagelkerke R^2 for each of these models

```{r}

fmsb::NagelkerkeR2(mdl\_stepAIC\_eilts\_pruned)

fmsb::NagelkerkeR2(mdl\_stepAIC\_eilts\_grown)

fmsb::NagelkerkeR2(mdl\_stepAIC\_eilts\_both)

```