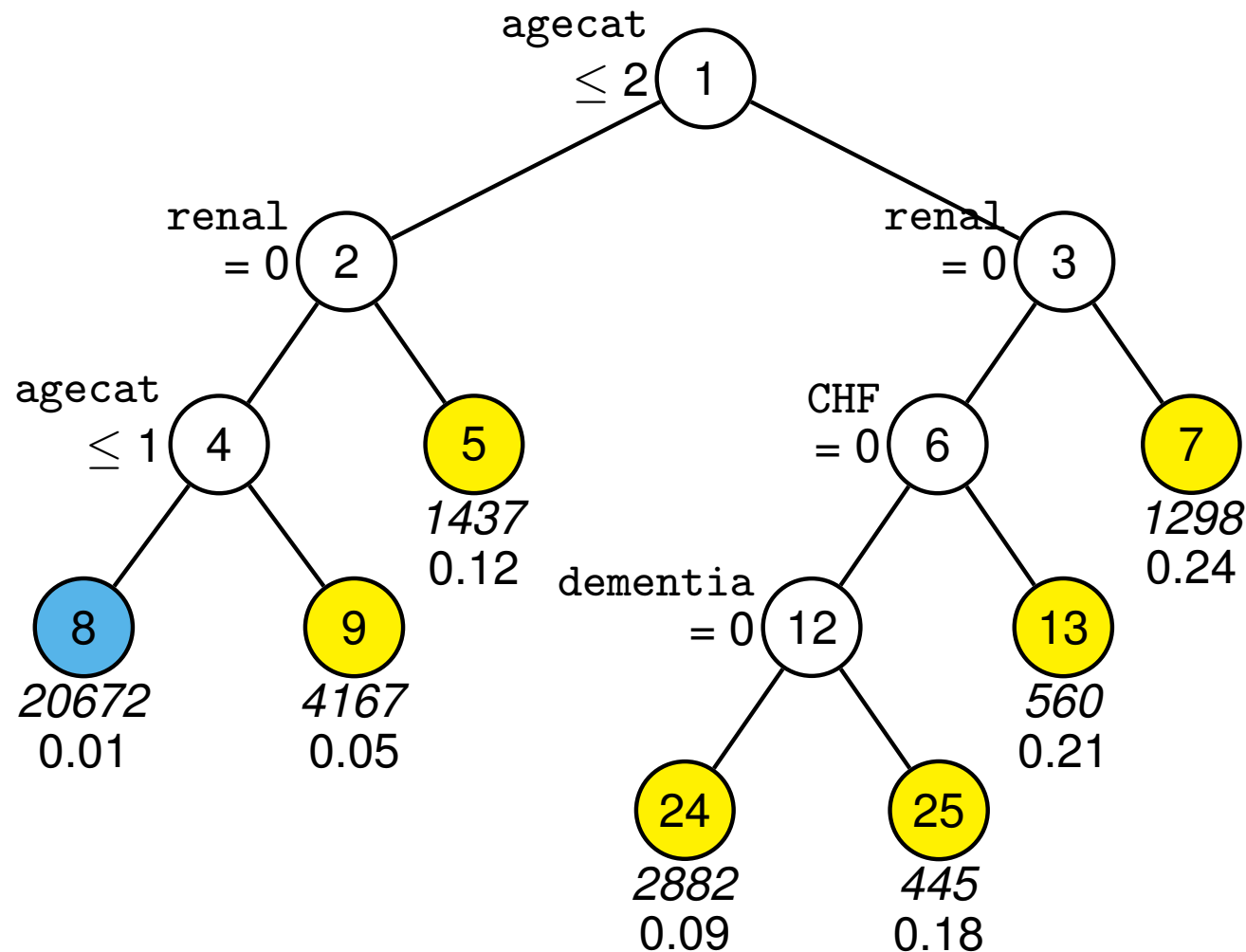
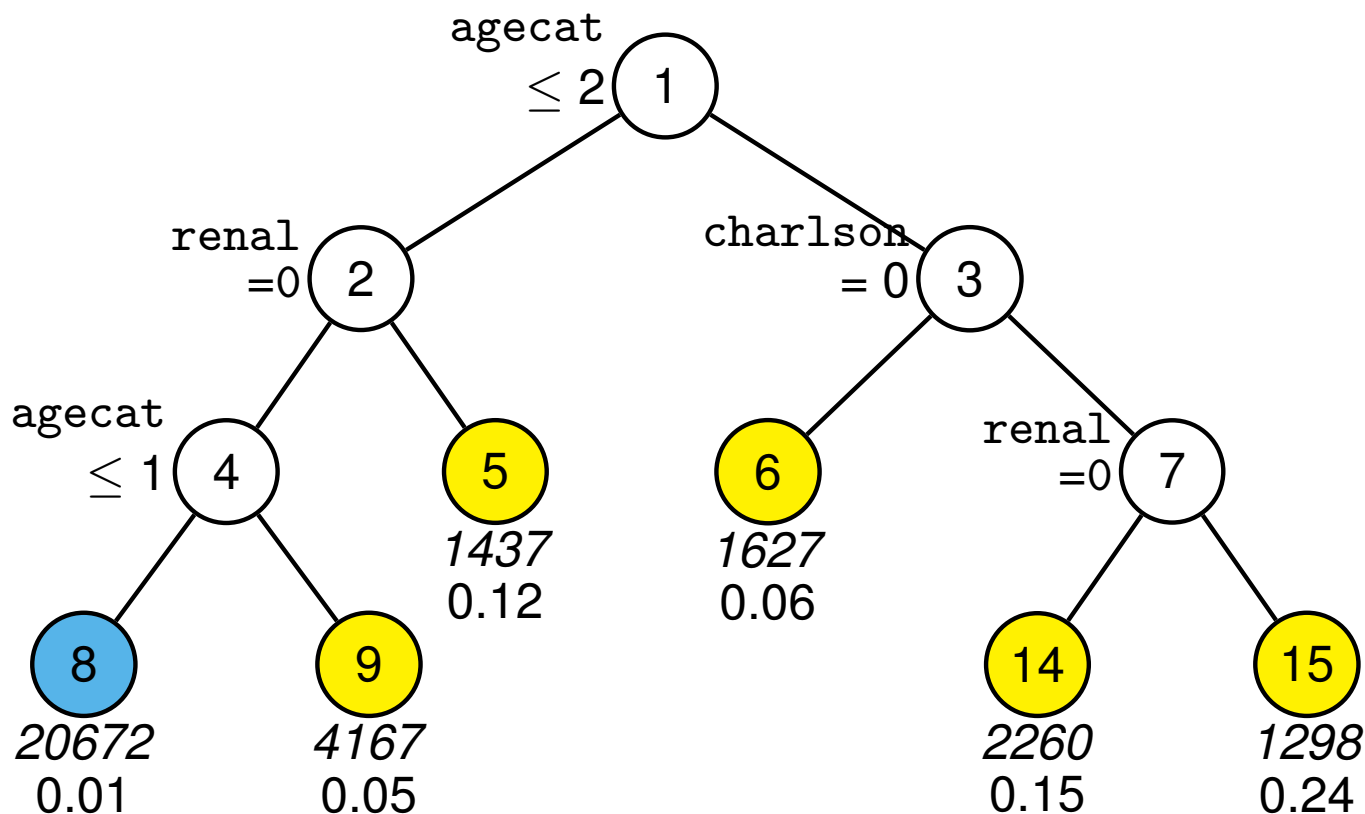


Regression tree (31461 obs, without charlson)



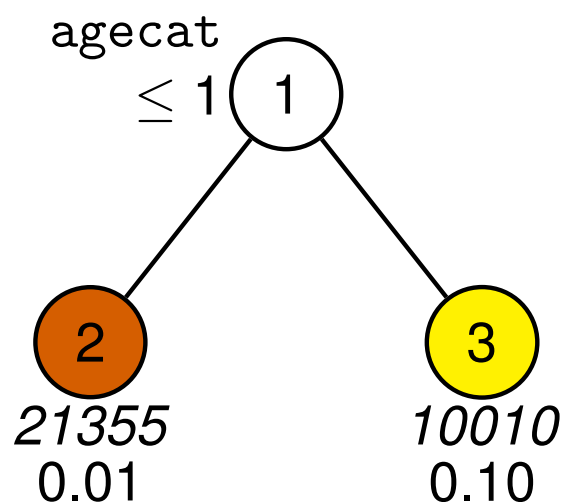
- Sample size (*in italics*) and mortality rate printed below nodes
- Terminal nodes with mortality rates above and below value of 0.04 at root node are colored yellow and skyblue, respectively

Regression tree (31461 obs, with charlson)



- Sample size (*in italics*) and mortality rate printed below nodes
- Terminal nodes with mortality rates above and below value of 0.04 at root node are colored yellow and skyblue, respectively

Logistic regression tree (without charlson and American Indian & Alaska Native)

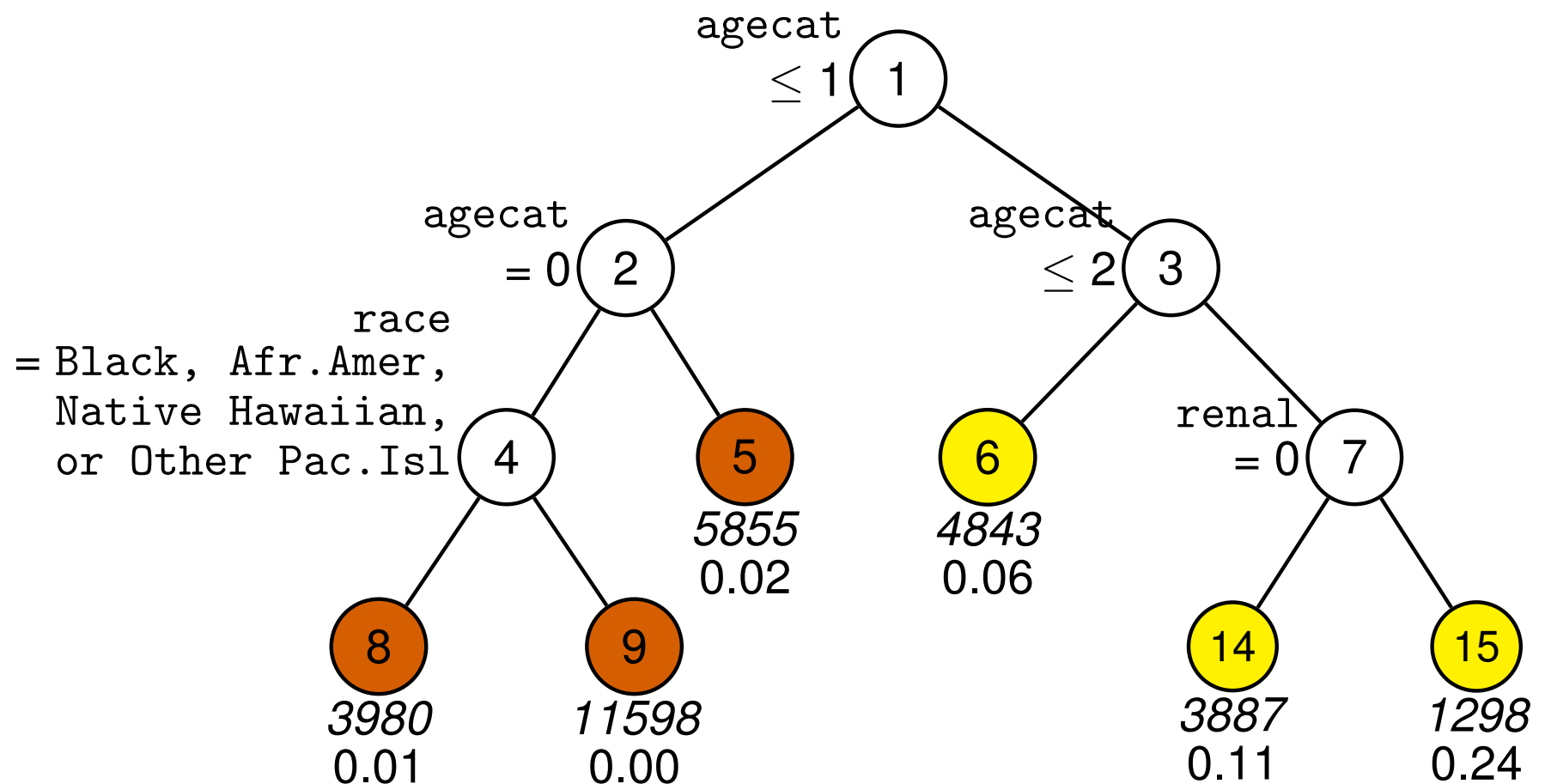


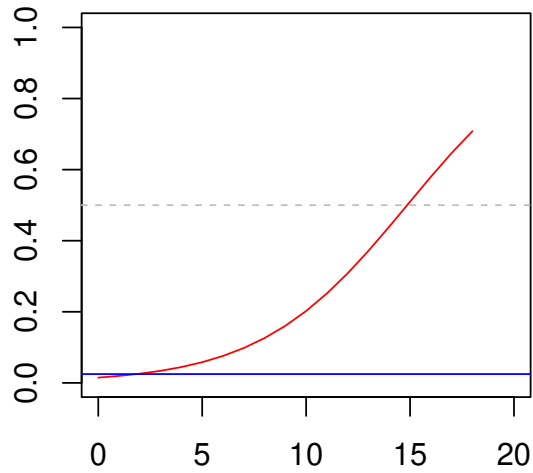
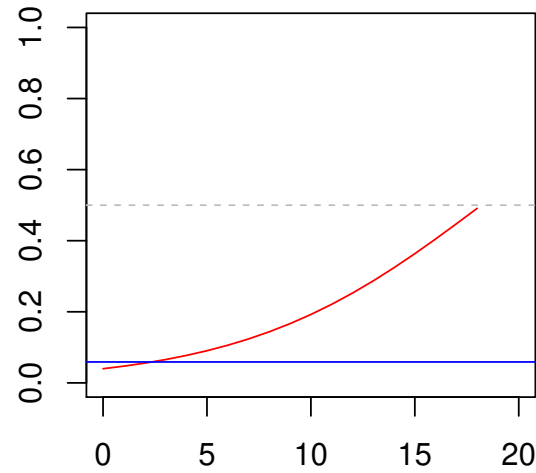
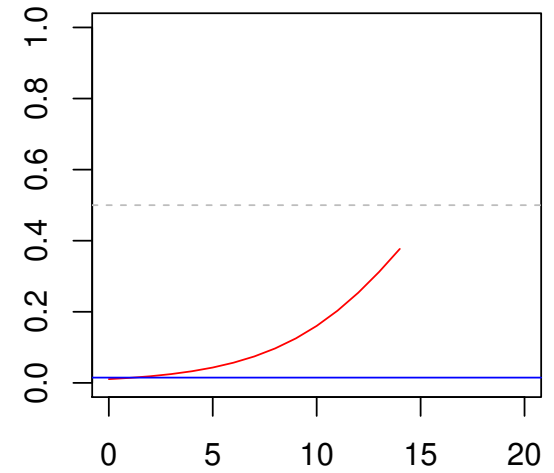
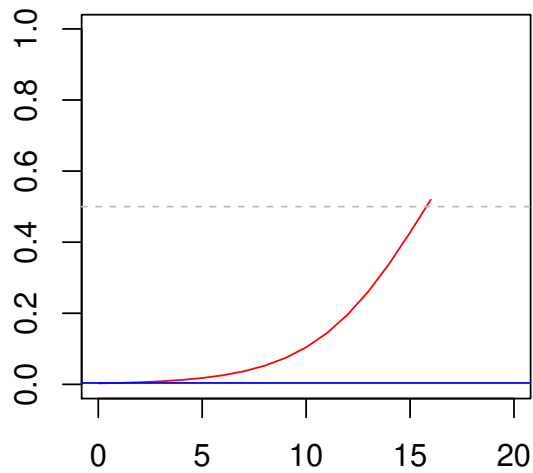
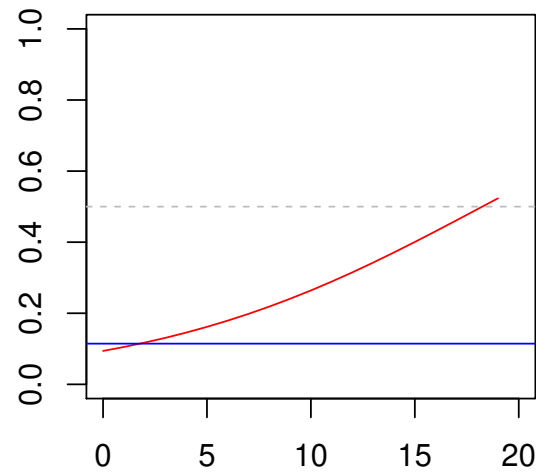
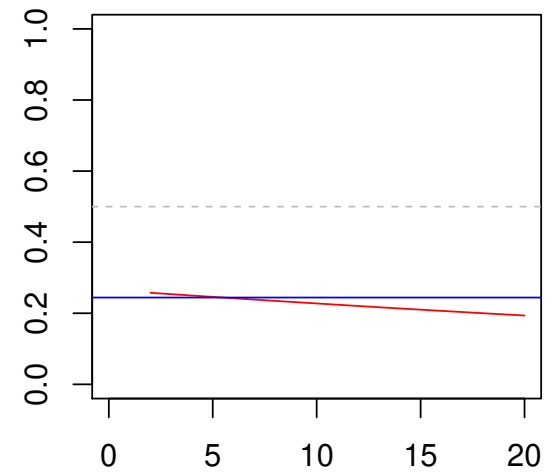
- At each split, an observation goes to the left branch if and only if the condition is satisfied
- Sample size (*in italics*) and estimated probability of death below nodes
- Logistic regression model fitted to each node

| | agecat ≤ 1 | | agecat > 1 | |
|-------------|-----------------|---------|--------------|---------|
| | Coef | P-value | Coef | P-value |
| (Intercept) | -6.626 | 0.000 | -4.760 | 0.000 |
| renal | 1.192 | 0.000 | 0.686 | 0.000 |
| agecat | 0.828 | 0.248 | 0.492 | 0.000 |
| CHF | 0.470 | 0.034 | 0.357 | 0.000 |
| MI | 0.940 | 0.000 | 0.608 | 0.000 |
| PVD | 0.128 | 0.630 | -0.103 | 0.288 |
| cerebro | 0.101 | 0.709 | 0.071 | 0.447 |
| dementia | 0.550 | 0.425 | 0.339 | 0.001 |
| diabetes | 0.233 | 0.166 | 0.036 | 0.653 |
| cancer | 0.112 | 0.713 | -0.129 | 0.219 |
| CPD | 0.293 | 0.069 | 0.192 | 0.017 |
| mildliver | 0.605 | 0.005 | -0.009 | 0.949 |
| modsevliv | 1.488 | 0.000 | 0.546 | 0.135 |

| | agecat ≤ 1 | | agecat > 1 | |
|---------------------------------------|-----------------|---------|--------------|---------|
| | Coef | P-value | Coef | P-value |
| sex.M | 0.817 | 0.000 | 0.463 | 0.000 |
| race.Black or African American | 1.478 | 0.036 | 0.733 | 0.011 |
| race.Native Hawaiian or Other Pacific | 1.248 | 0.310 | 2.020 | 0.000 |
| race.Unknown | 0.398 | 0.580 | 0.204 | 0.496 |
| race.White | 0.727 | 0.305 | 0.457 | 0.110 |
| metastatic | 0.881 | 0.080 | 0.479 | 0.014 |
| hemipara | 0.523 | 0.211 | -0.513 | 0.016 |
| RD | 0.135 | 0.719 | 0.115 | 0.514 |
| PUD | -0.088 | 0.835 | -0.396 | 0.067 |
| aids | 0.527 | 0.174 | 0.212 | 0.599 |

Logistic regression tree (31,461 obs, with charlson as sole linear predictor)



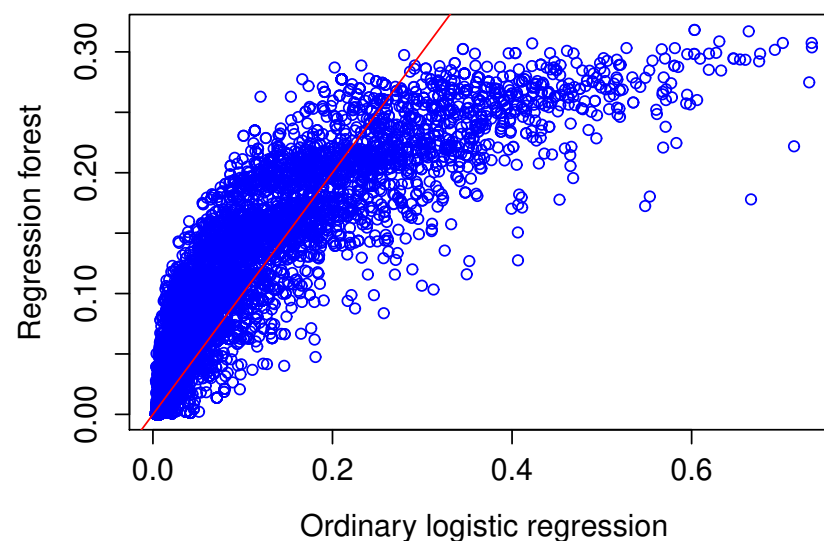
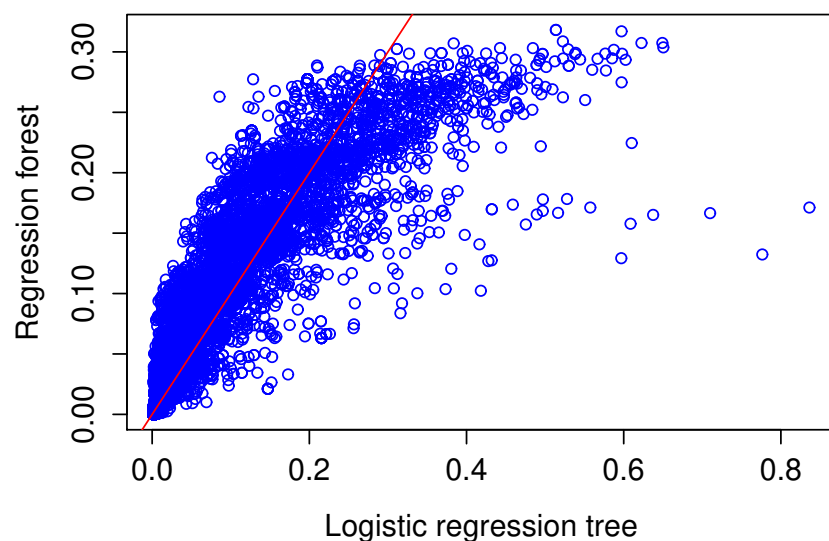
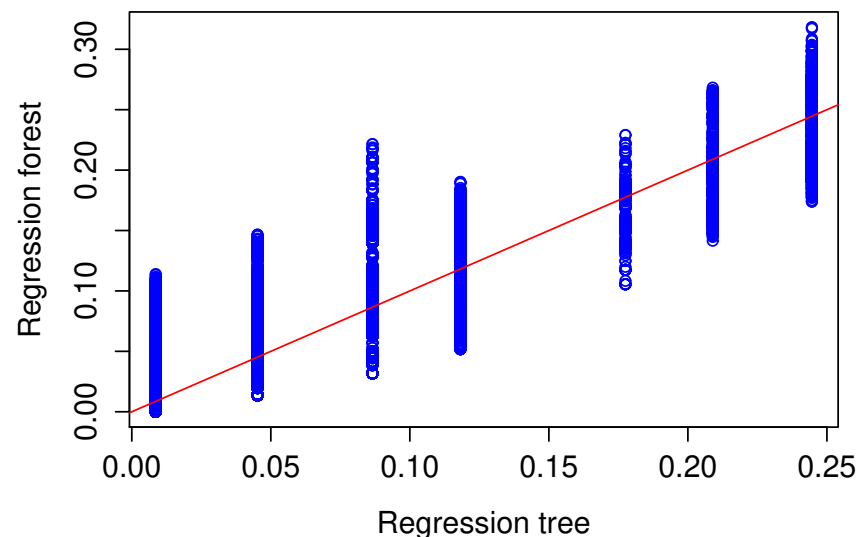
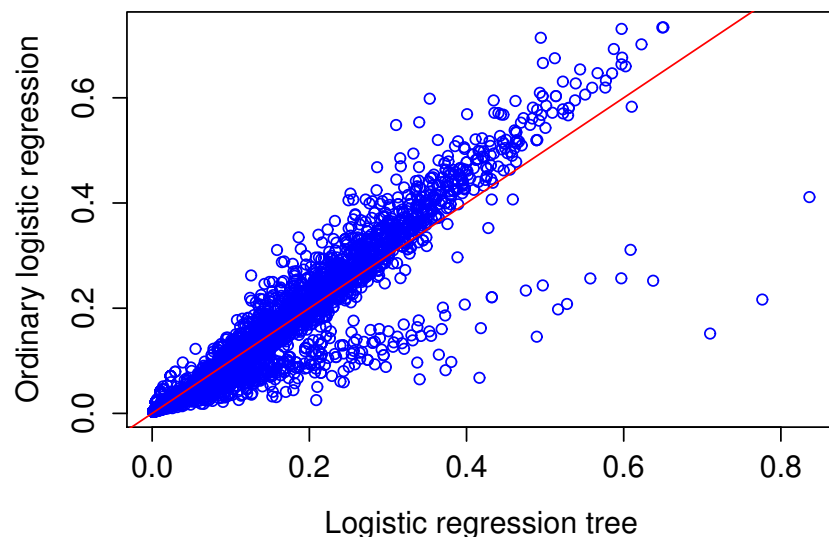
Node 5**Node 6****Node 8****Node 9****Node 14****Node 15**

Red line = fitted $P(\text{died})$; blue line = observed death rate

GUIDE models

- Regression tree with/without American Indian & Alaska Native and with/without charlson
- Regression forest with/without American Indian & Alaska Native and with/without charlson
- Logistic regression tree without American Indian & Alaska Native and with/without charlson

P(died) w/o charlson & Am. Indian & Alaska Nat.



Accuracy vs interpretability

Most
accurate

Most
interpretable

Classification
or regression
forest

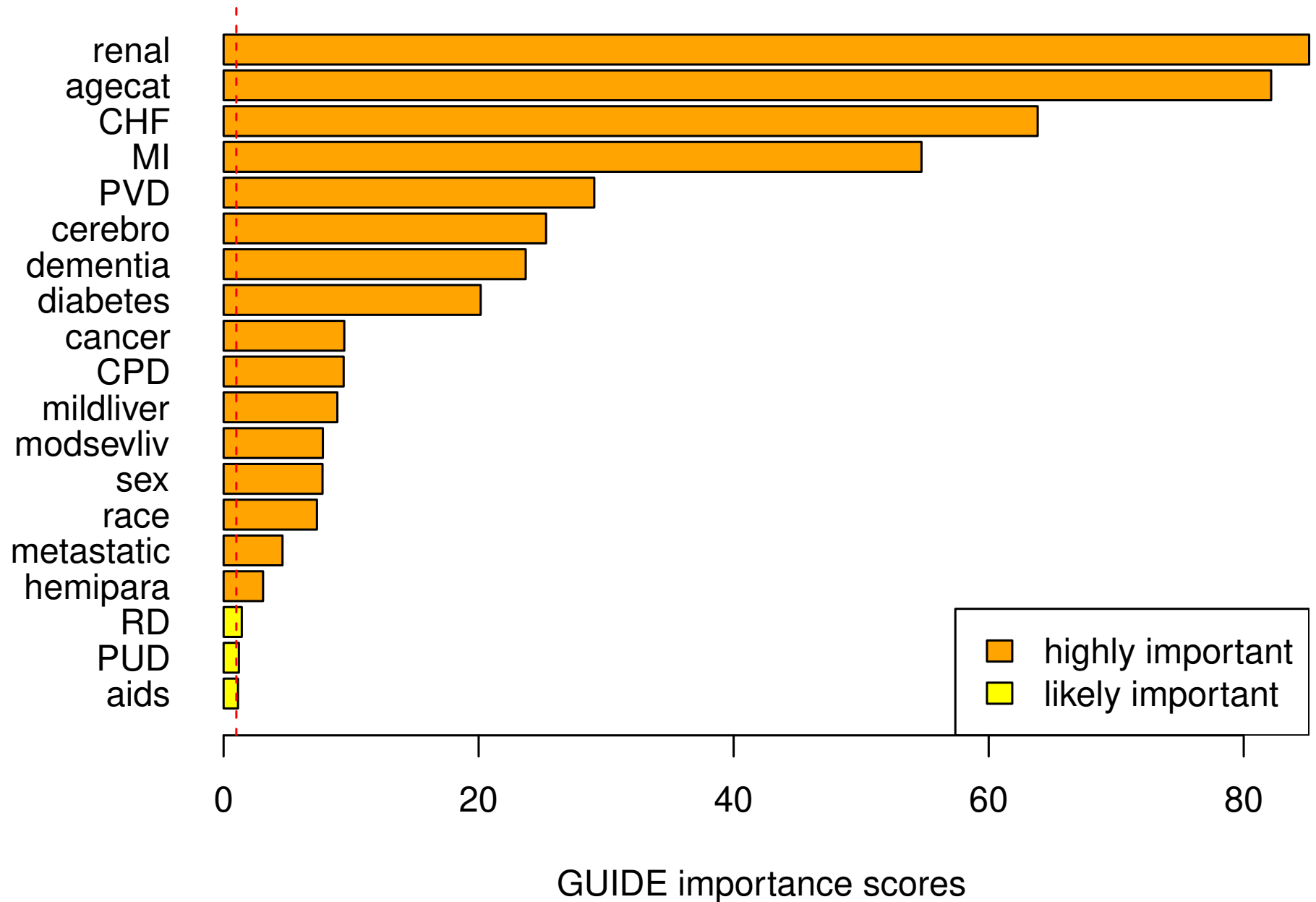
Logistic
regression
tree

Piecewise
constant
tree

Logistic models

- Ordinary logistic regression does not allow `charlson` and `American Indian & Alaska Native`
- Logistic regression tree does not allow `American Indian & Alaska Native`

GUIDE importance scores



About GUIDE

- GUIDE algorithm and software have been in development for 30+ years
- GUIDE manual and free compiled code for Linux, Mac OS X and Windows are available at www.stat.wisc.edu/~loh/guide.html
- GUIDE is not implemented in R but can be used in R (see manual)
- Key references: Loh and Vanichsetakul (1988), Chaudhuri et al. (1994, 1995), Loh and Shih (1997), Kim and Loh (2001), Loh (2002, 2009, 2014, 2019), Loh and Zheng (2013), and Loh et al. (2015, 2016, 2019a,b)

Things to do before next Tuesday

1. Go to <http://pages.stat.wisc.edu/~loh/guide.html> and install GUIDE
2. Read the GUIDE manual
<http://www.stat.wisc.edu/~loh/treeprogs/guide/guideman.pdf>
3. Read article: Loh, W.-Y. (2011), Classification and regression trees, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol.1, 14–23
<http://www.stat.wisc.edu/~loh/treeprogs/guide/wires11.pdf>