Comparison of models for estimating age at motor onset in HD

Peter Holmans[1] and Oliver Didcote[2]

1. Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, United Kingdom
2. School of Medicine, Cardiff University, United Kingdom

Email: HolmansPA@cardiff.ac.uk

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Introduction

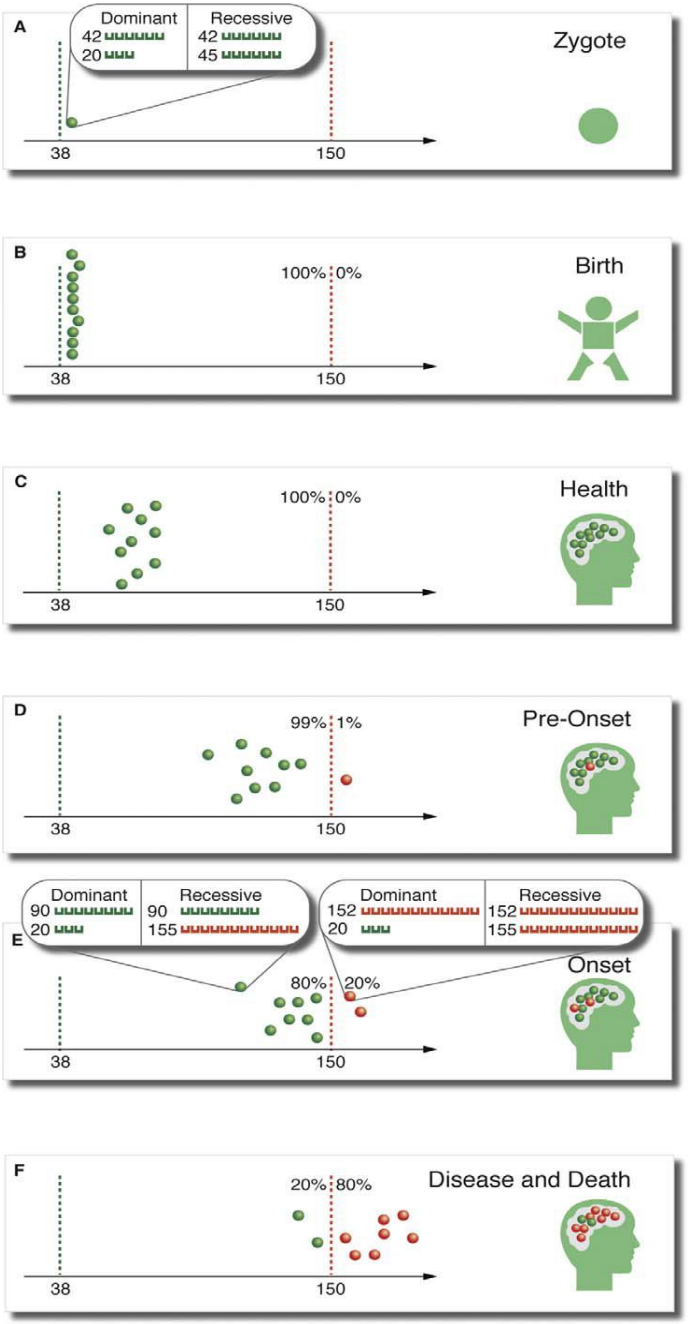
- Increased CAG length is associated with earlier age at onset (AAO) in HD, accounting for about 50-60% of variation (1)
- The power to detect other risk factors modifying AAO (for example, genetic) is increased by taking the effects of CAG length into account.
- It is therefore important to model the effect of CAG length on AAO as accurately as possible, over as wide a range of CAG length as possible.
- We use the Enroll PDS5 dataset to compare the AAO prediction of the commonly used Langbehn model (2) to that of the model proposed by Kaplan et al. (3)

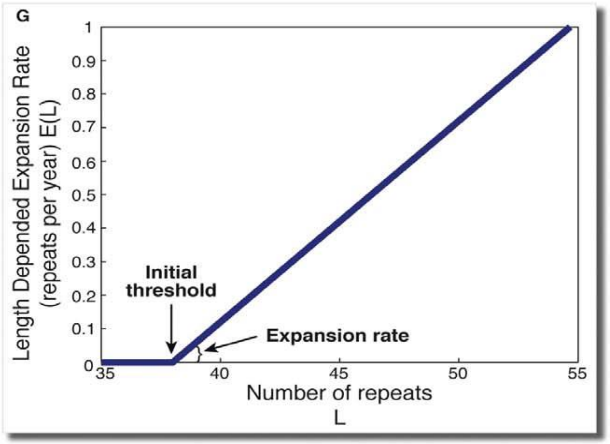
Langbehn model

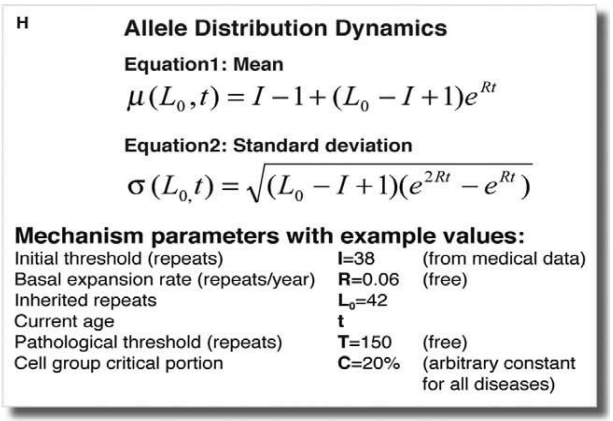
- AAO = 21.54 + Exp(9.556 – 0.1460CAG)
- Assumes a linear relationship between CAG length and ln(AAO)
- This fits well for CAG=40-55 but less well for CAG outside this range
- Does not suggest a mechanism by which CAG influences AAO

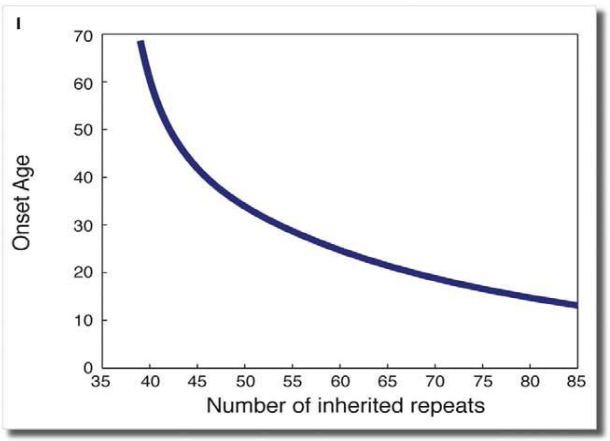
Kaplan Model

(A) Patient inherits a trinucleotide repeat that exceeds disease-specific threshold (green line)
(B) Repeat lengths in each cell are initially clustered around the inherited value.
(C-E) Repeat lengths increase stochastically. Disease onset occurs when a sufficient proportion (here, 20%) cross a pathological threshold (red line)
(F) The disease progresses toward death as more cells cross the target threshold.
(G) The rate of allele expansion E is a linear function of the number of repeats above the initial threshold.
(H) Equations for the mean and standard deviation of allele size as a function of the patient's age t, inherited number of repeats L_0, and the mechanism parameters.
(I) The mechanism predicts an exponentially decreasing onset curve similar to curves obtained from clinical data for trinucleotide diseases







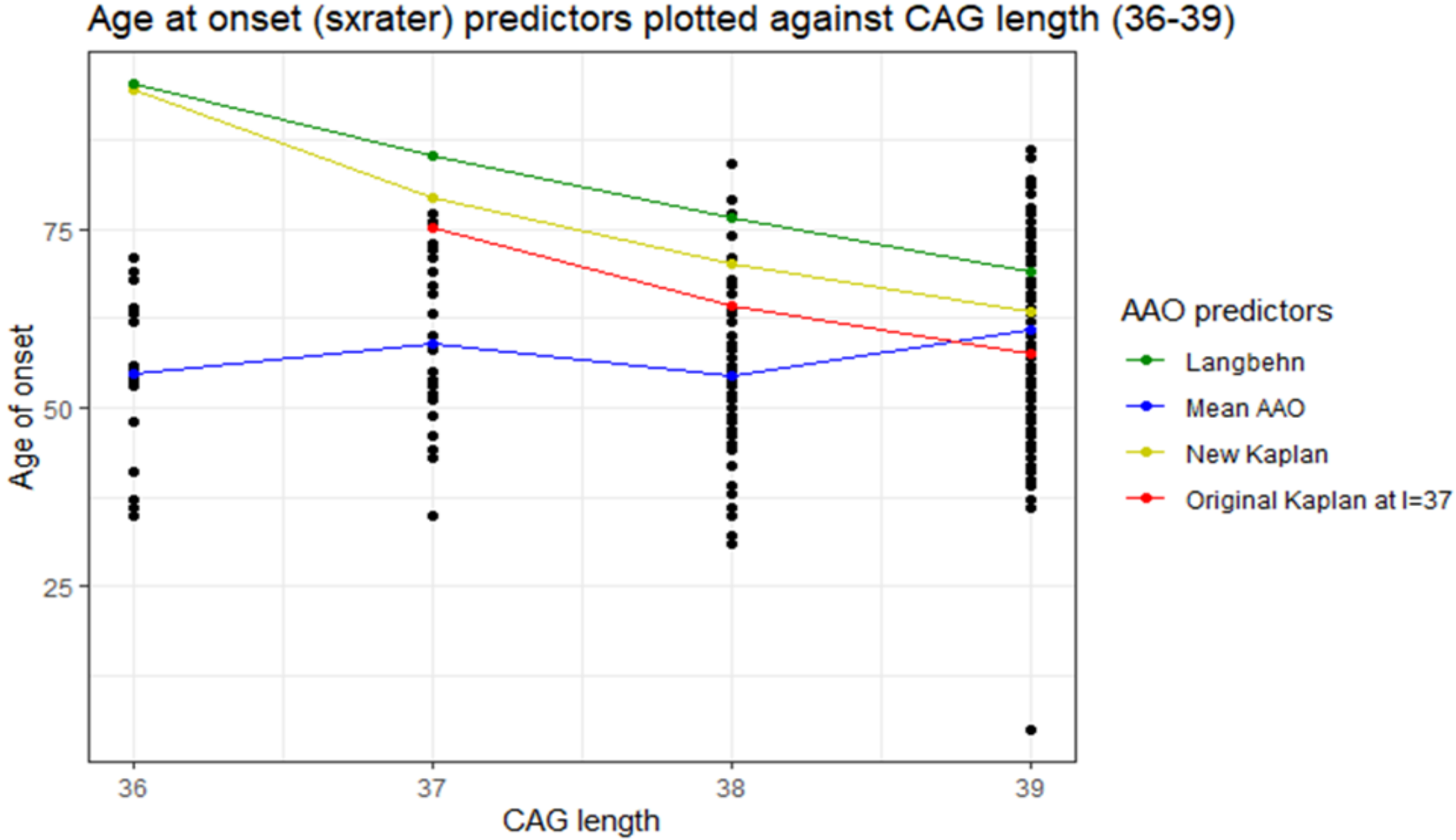


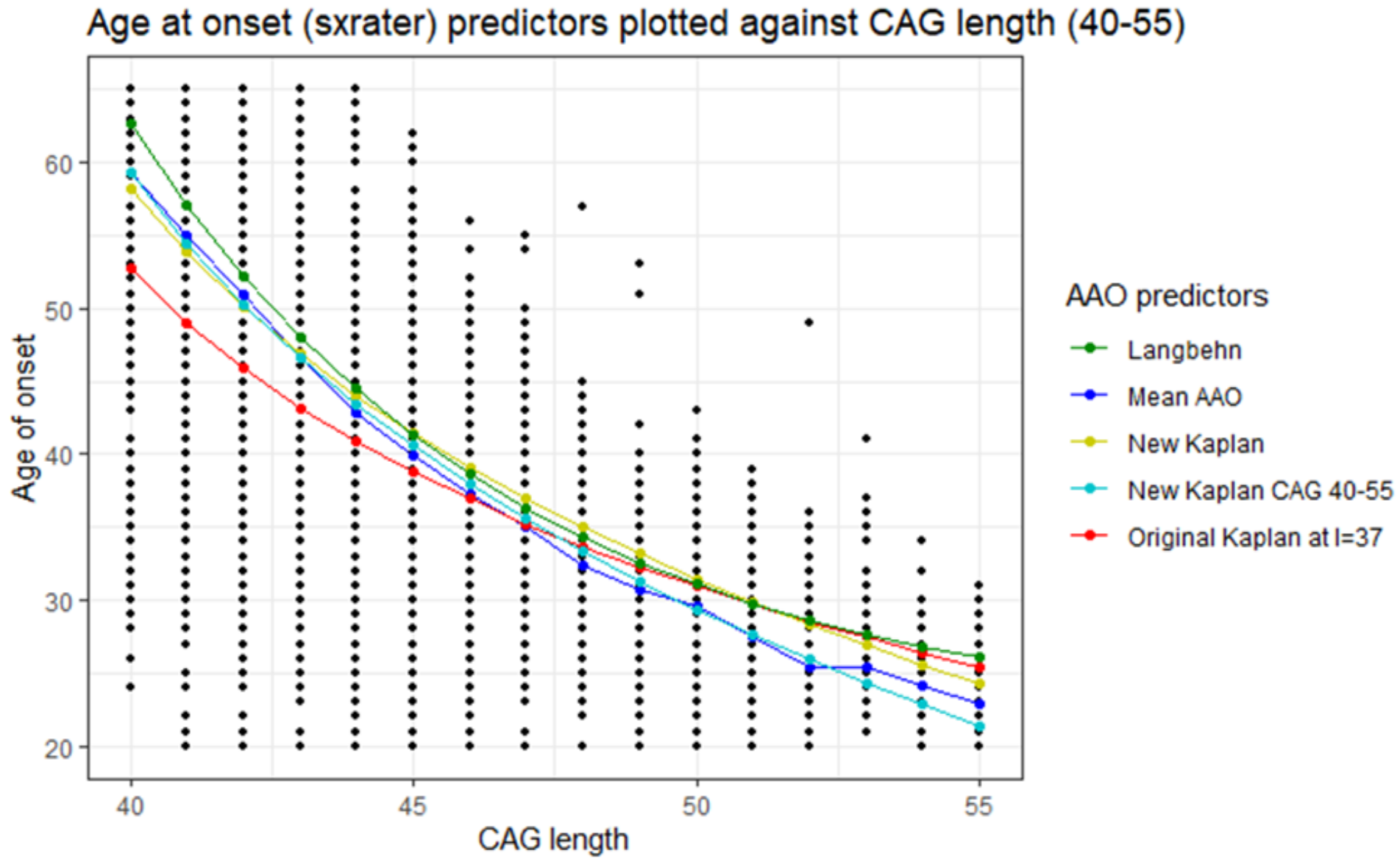
Fitting the models to the data

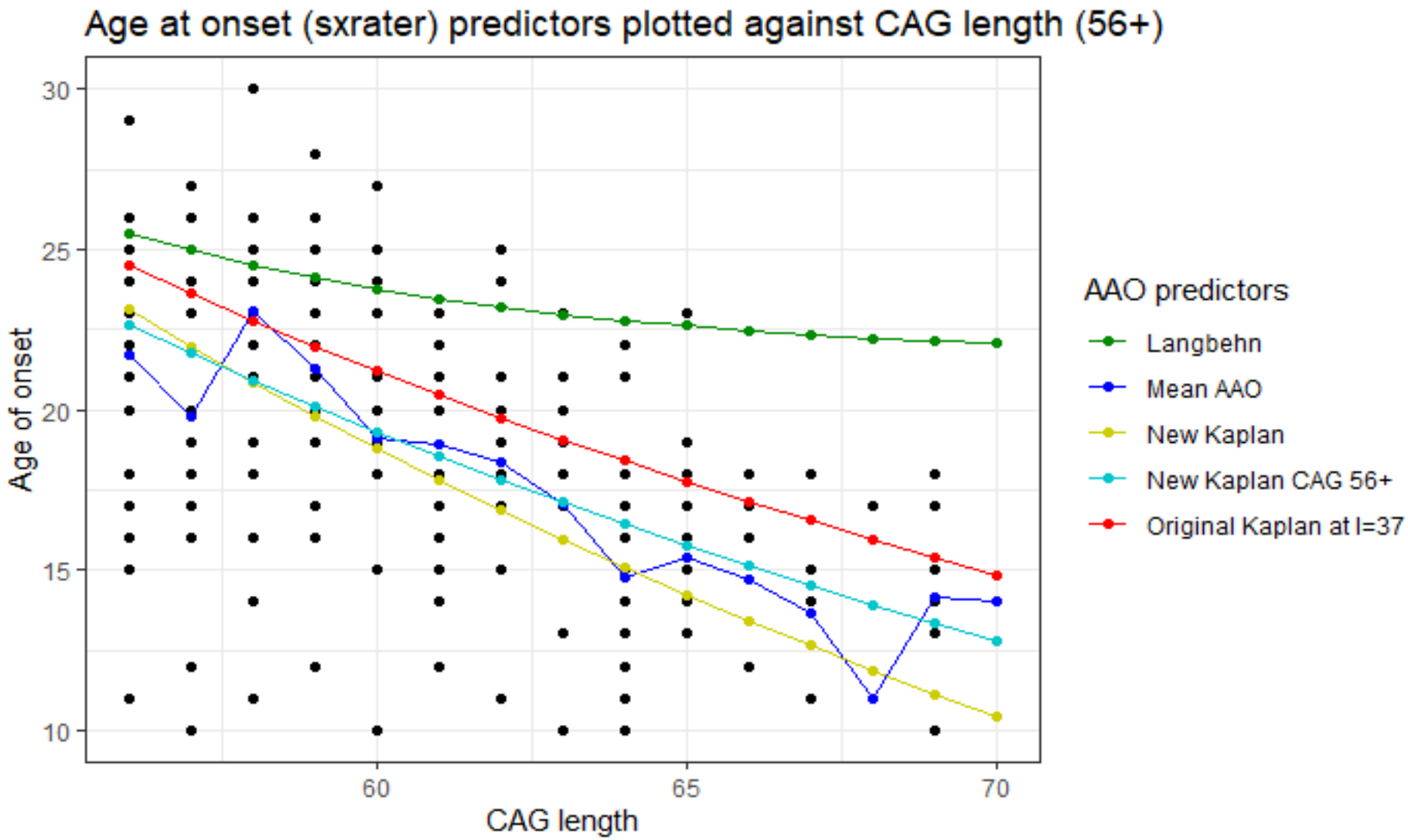
- 10,929 individuals with sxrater estimate of onset (results similar if ccmtrage used instead)
- CAG range = 36-70
- Fit models to all CAG, 36-39, 40-55, 56+
- Measure of AAO prediction accuracy:
  $R^2 = 1- [\Sigma (AAO – Predicted\ AAO)^2 / \Sigma (AAO – mean(AAO))^2]$
- Grid search of values of pathogenic threshold (T) and CAG expansion rate (R) in Kaplan model
- Repeat threshold for disease: I=36 (also I=37 used by Kaplan)

Overall model prediction (R^2)

Model CAG range	T (CI)	R (CI)	Test CAG range	N	R^2 (Kaplan)	R^2(Langbehn)
36+(ALL)	90 (75-150)	0.036 (0.030-0.052)	36+	10929	0.555	0.520
			36-39	319	-1.26	-2.11
			40-55	10398	0.550	0.540
			56+	212	0.157	-0.516
40-55	80 (75-115)	0.032 (0.030-0.044)	40-55	10398	0.561	0.540
56+	105 (90-250)	0.047 (0.034-0.105)	56+	212	0.180	-0.516


Age at onset (sxrater) predictors plotted against CAG length (36-39)
Age of onset
75
50
25
36
37
38
39
CAG length
AAO predictors
Langbehn
Mean AAO
New Kaplan
Original Kaplan at I=37


Age at onset (sxrater) predictors plotted against CAG length (40-55)
Age of onset
60
50
40
30
20
40
45
50
55
CAG length
AAO predictors
Langbehn
Mean AAO
New Kaplan
New Kaplan CAG 40-55
Original Kaplan at l=37


Age at onset (sxrater) predictors plotted against CAG length (56+)
Age of onset
30
25
20
15
10
60
65
70
CAG length
AAO predictors
Langbehn
Mean AAO
New Kaplan
New Kaplan CAG 56+
Original Kaplan at l=37

Conclusions

- The Kaplan and Langbehn models give similar accuracy in predicting AAO for CAG=40-55, with Kaplan slightly more accurate for CAG>50.
- The Kaplan model is more accurate than the Langbehn model for CAG>56, although accuracy is reduced compared to CAG=40-55
- Neither model could predict AAO for CAG=36-39, with both models overestimating AAO.
- Using the Kaplan model to predict AAO could enable the inclusion of people with CAG>56 in GWAS.

References

1. Paulsen et al. 2008 J. Neurol. Neurosurg. Psychiatry 79, 874–880
2. Langbehn et al. 2004 Clin Genet 65(4):267–277
3. Kaplan et al. 2007, PLoS Comp Biol 3(11): e235