

#HackEbola

Fundamentals of infectious disease forecasting

a statistical perspective

Nicholas Reich
HackEbola @ UMass-Amherst
21 November 2014

Infectious disease forecasting

biological factors

- infectiousness
- virulence
- incubation period
- pathogen interactions
- climate/weather
- disease vectors



behavioral factors

- travel patterns
- social contacts

Very few “successful”
real-time infectious
disease forecasting efforts.

Most high-profile work right now is done by
Jeff Shaman at Columbia on flu and Ebola.

evaluation

evaluation

evaluation

fit your models

training
dataset

choose a model

validation
dataset

apply, evaluate

testing
dataset

fit your models

training
dataset

choose a model

validation
dataset

apply, evaluate

testing
dataset

If you don't follow this recipe,
you will "overfit" your data and
predictions will be garbage.

fit your models

training
dataset

choose a model

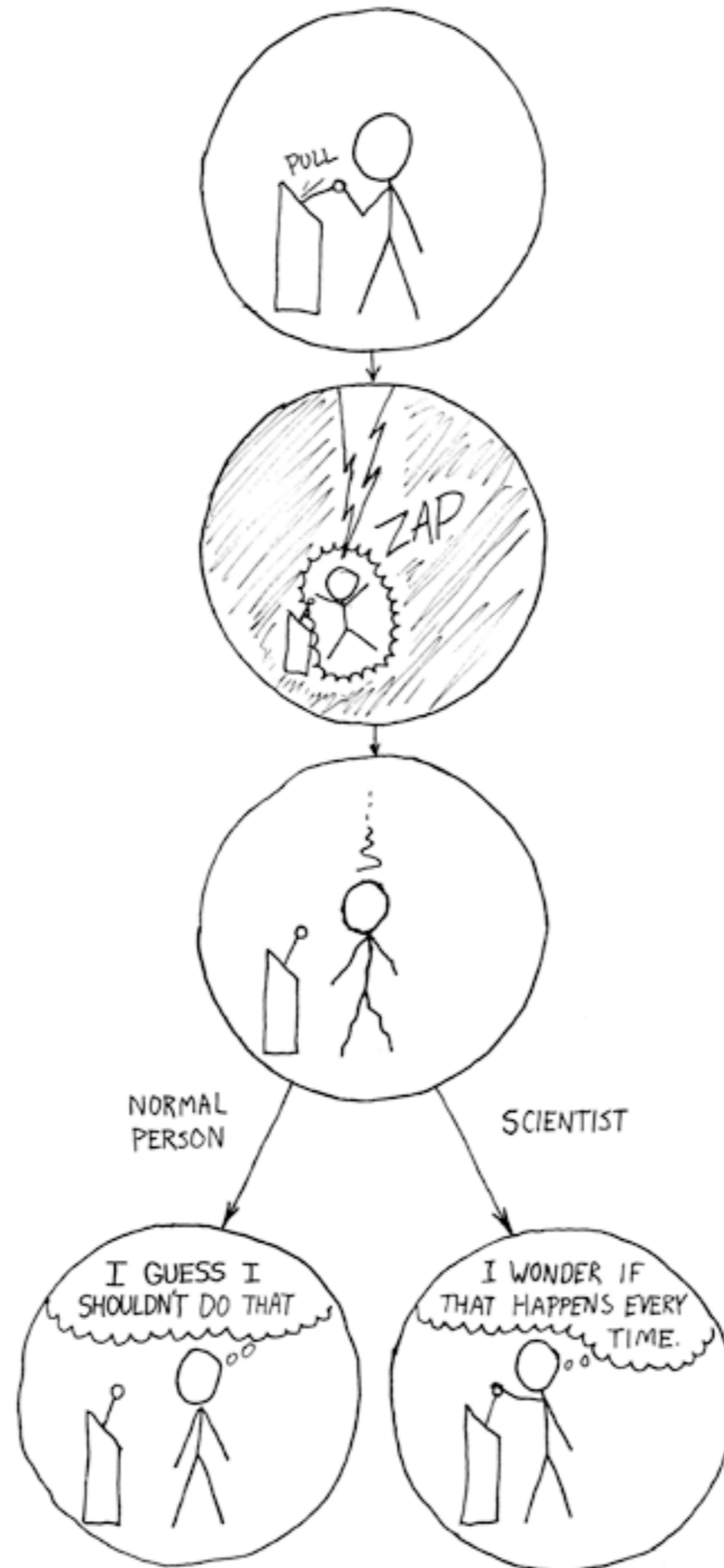
validation
dataset

apply, evaluate

testing
dataset

If you don't follow this recipe,
you will "overfit" your data and
predictions will be garbage.

They might be still garbage.



validating a prediction!



Do your predictions
provide added value?

DETOUR



Do Nate Silver's electoral
predictions provide
added value?



Predicted number of senate seats, 2014 US Election

predicted

16 Democratic

19 Republican

actual

12 Democratic

23 Republican

4
off

$31/35 = 89\%$ accuracy!

Predicted number of senate seats, 2014 US Election

predicted

16 Democratic

19 Republican

actual

12 Democratic

23 Republican

4
off

$31/35 = 89\%$ accuracy!

But, these are my predictions based on whether Obama won the state in the 2012 election. This is not from Nate Silver.

Predicted number of senate seats, 2014 US Election

predicted

14 Democratic

21 Republican

actual

12 Democratic

23 Republican

2

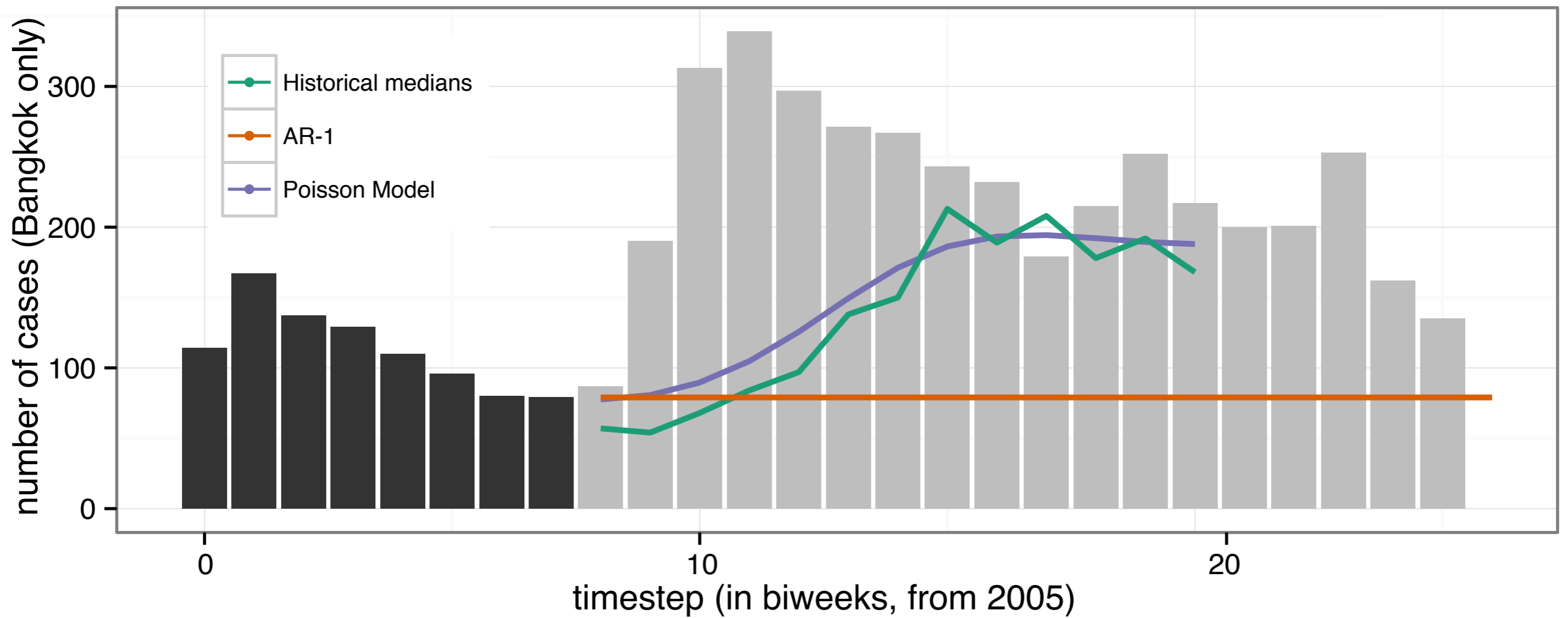
$33/35 = 94\%$ accuracy!

Arguably, not much better than my no-fuss predictions...

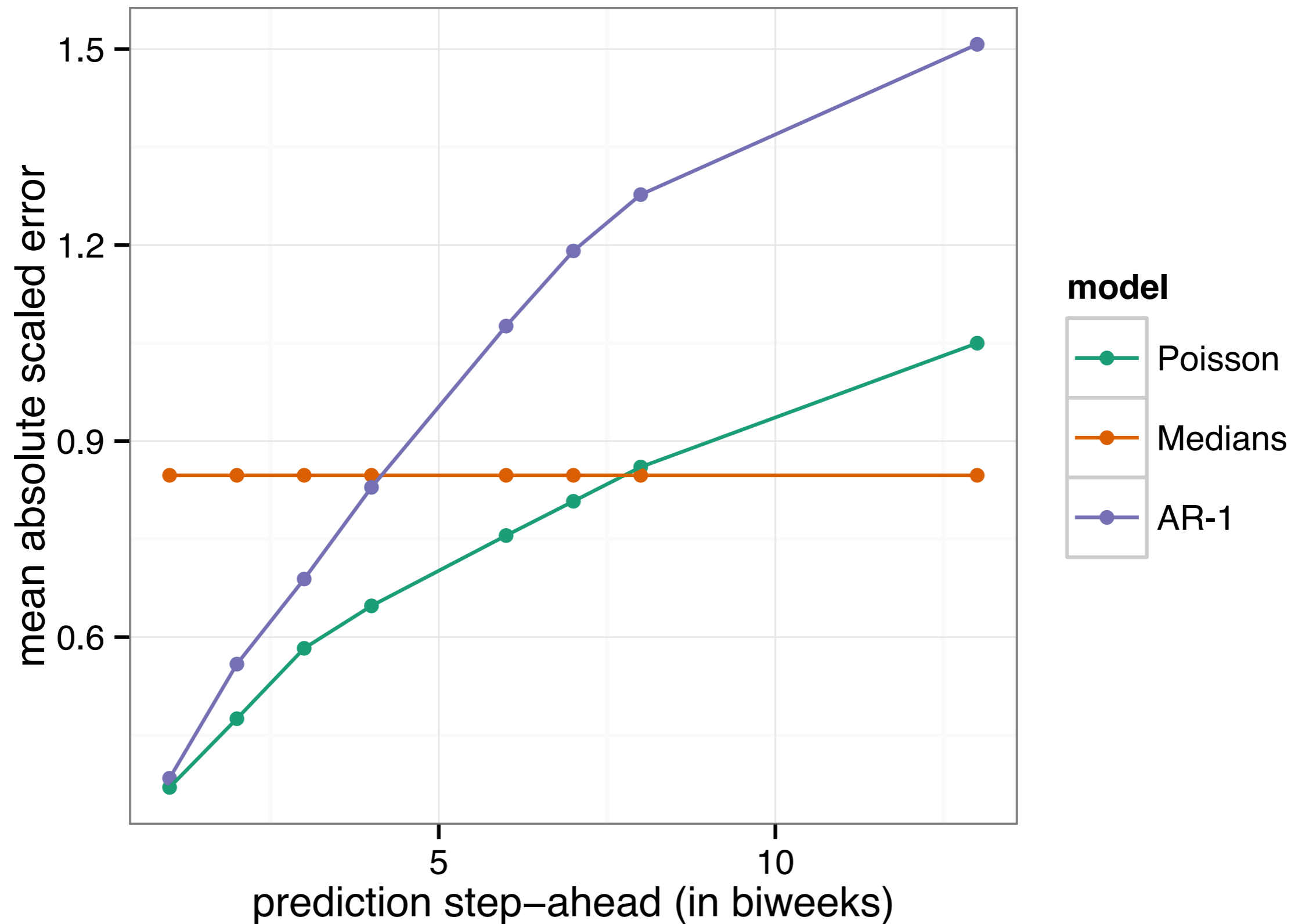
Back to
infectious disease



Three predictions of dengue fever cases in Bangkok, Thailand (2005 case data)



Especially for forecasts far into the future, important to compare to relevant reference models!



Open question:
what would be a good reference
forecast model for Ebola?

Some tools to consider, if you go the way of forecasts

- SEIR models of infectious disease (see, e.g., [Shaman et al. 2014](#))
- Poisson or Negative Binomial regression
- Statistical cross-validation of fitted models
- For forecasts, **simple** is better!

HackEbola @ UMass

November 21-23, 2014

#HackEbola

Use challengepost.com site but don't be tied to their "Rules" page.

Focus on what you are interested in!

Possible projects

- Create and compare reference forecast models for Ebola.
- Create some forecasts.
- Create a tool to aggregate forecasts (now) and evaluate them down the road (hold people accountable!)
- Develop agent-based simulation model of Ebola.
- Create first responder maps (good project if you are intimidated by all the stats/coding).
- More ideas on challengepost.com