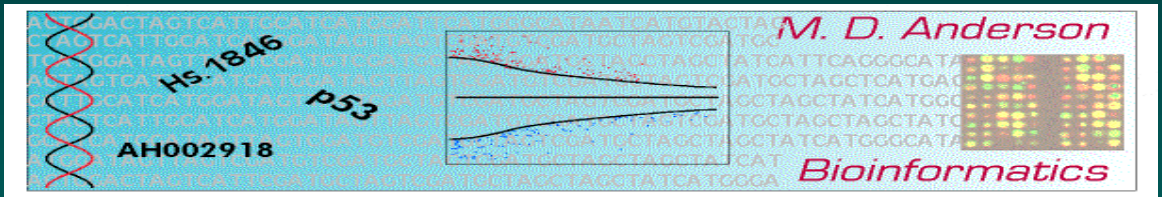
Vitamin D and Reproducibility

Keith A. Baggerly
Bioinformatics and Computational Biology
UT M. D. Anderson Cancer Center
kabagg@mdanderson.org

SISBID, July 17, 2017



Testing Reproducibility

This is an ongoing case study.

Last year, I was asked whether vitamin D supplementation might reduce preterm birth rates. Current recommendations make implementing high supplement interventions difficult.

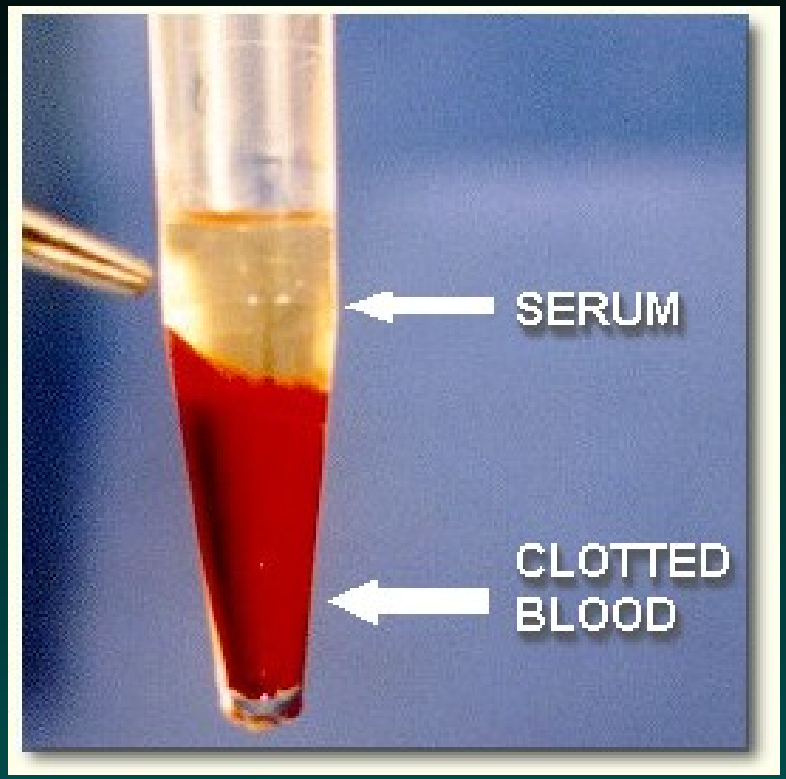
COI: the person asking was mom.

In addition to the preterm data, I reviewed the report in which the recommendations were derived.

This led to some questions.

As we dive into this, I want to try to keep our terms clear.

Nomenclature

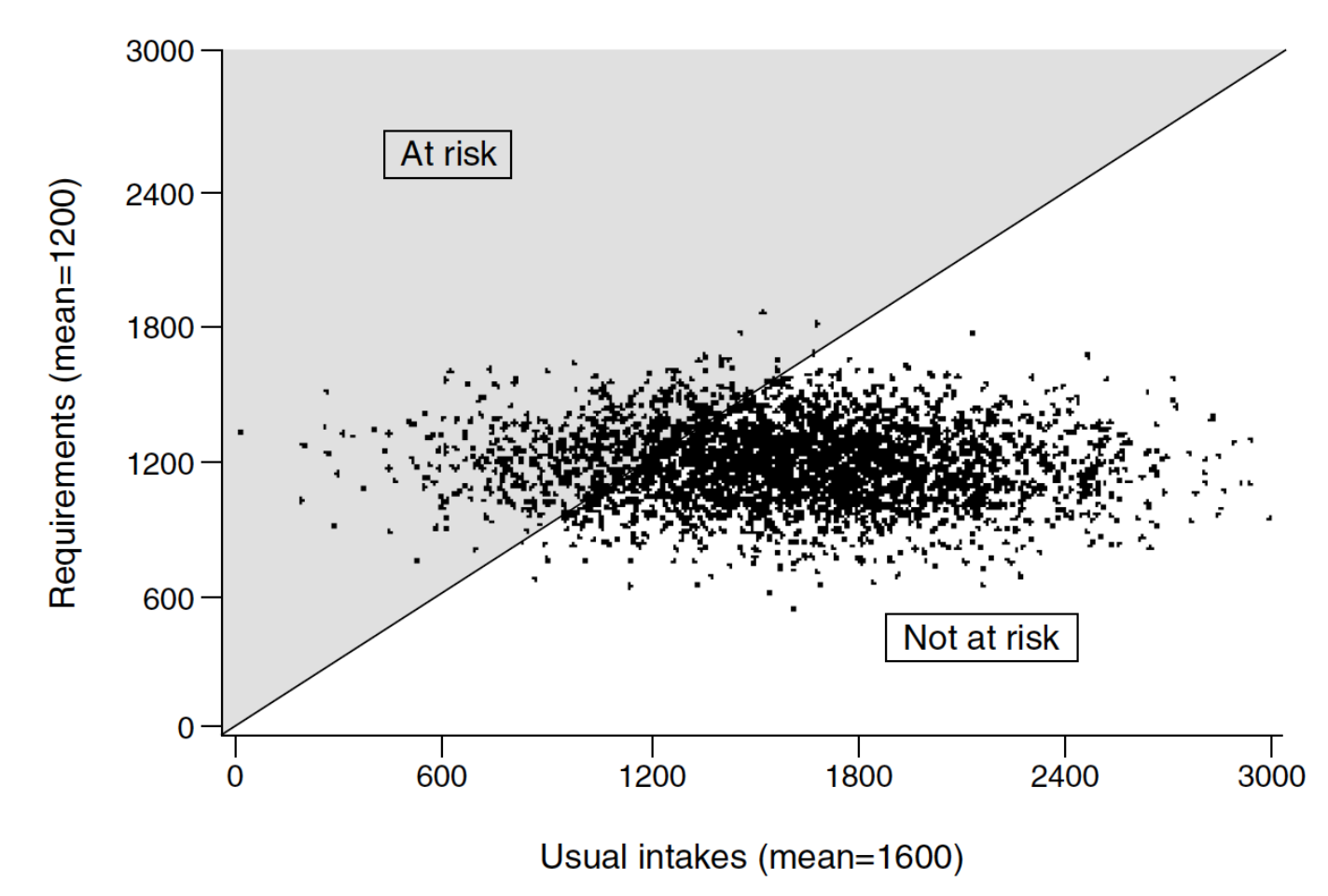


Intakes in IU

Serum Levels in ng/mL = 2.5 nmol/L

Requirements measured in terms of the latter

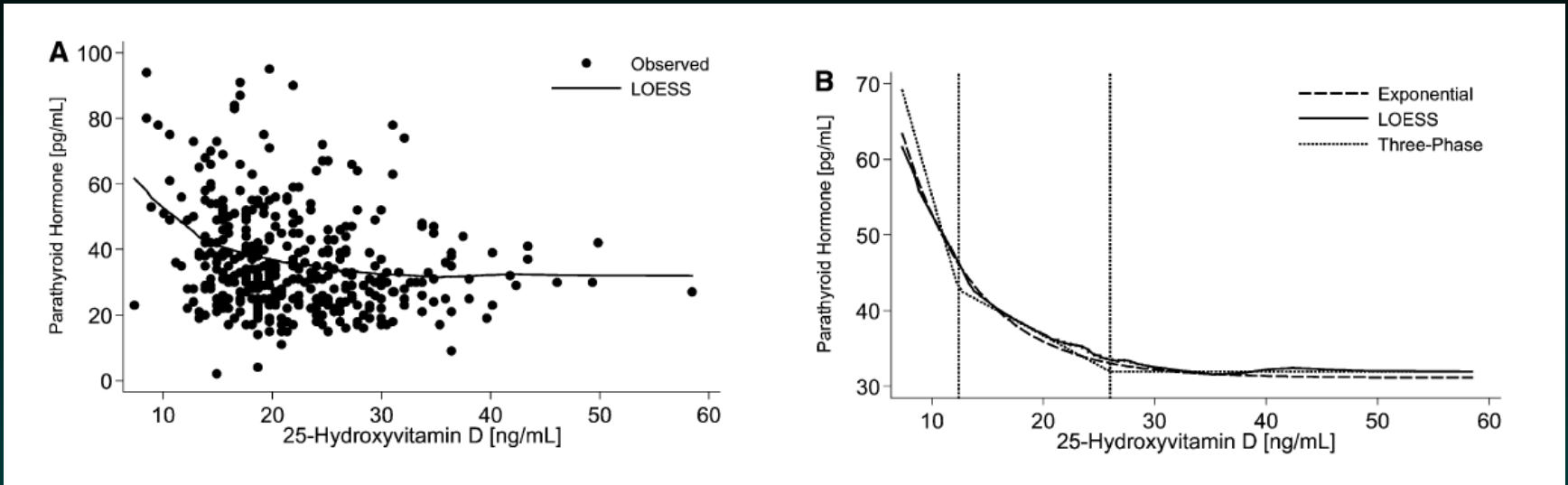
Modeling Intakes and Requirements



IOM 2000: DRIs in Dietary Assessment, Fig 4.2

Assume normality and model.

Defining Requirements in Other Units



Durazo-Arvizu et al, 2010, Fig 1

When is the nutrient product not biochemically limiting?

If vitamin D is too low to regulate calcium, parathyroid hormone (PTH) will increase and leach Ca from bones.

Requirements use serum levels (ng/mL); intakes use IU.

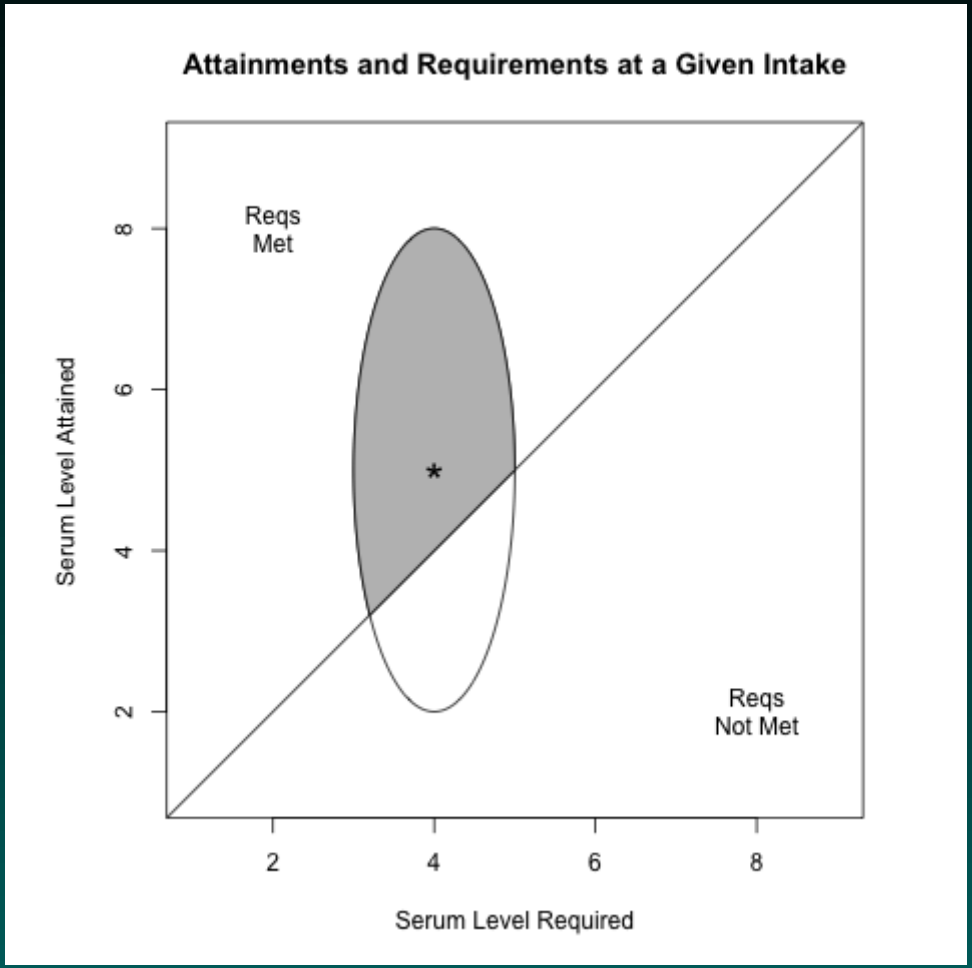
Requirements, Intakes, and Units

We need to fit a dose response curve relating intake levels to serum levels.

This curve is generally logarithmic in shape; you need to more than double the intake to double the serum level.

If there's marked variability in serum level attainments for a given intake, we need to model this new variation as well.

A Joint (Required, Attained) Model Given Intake



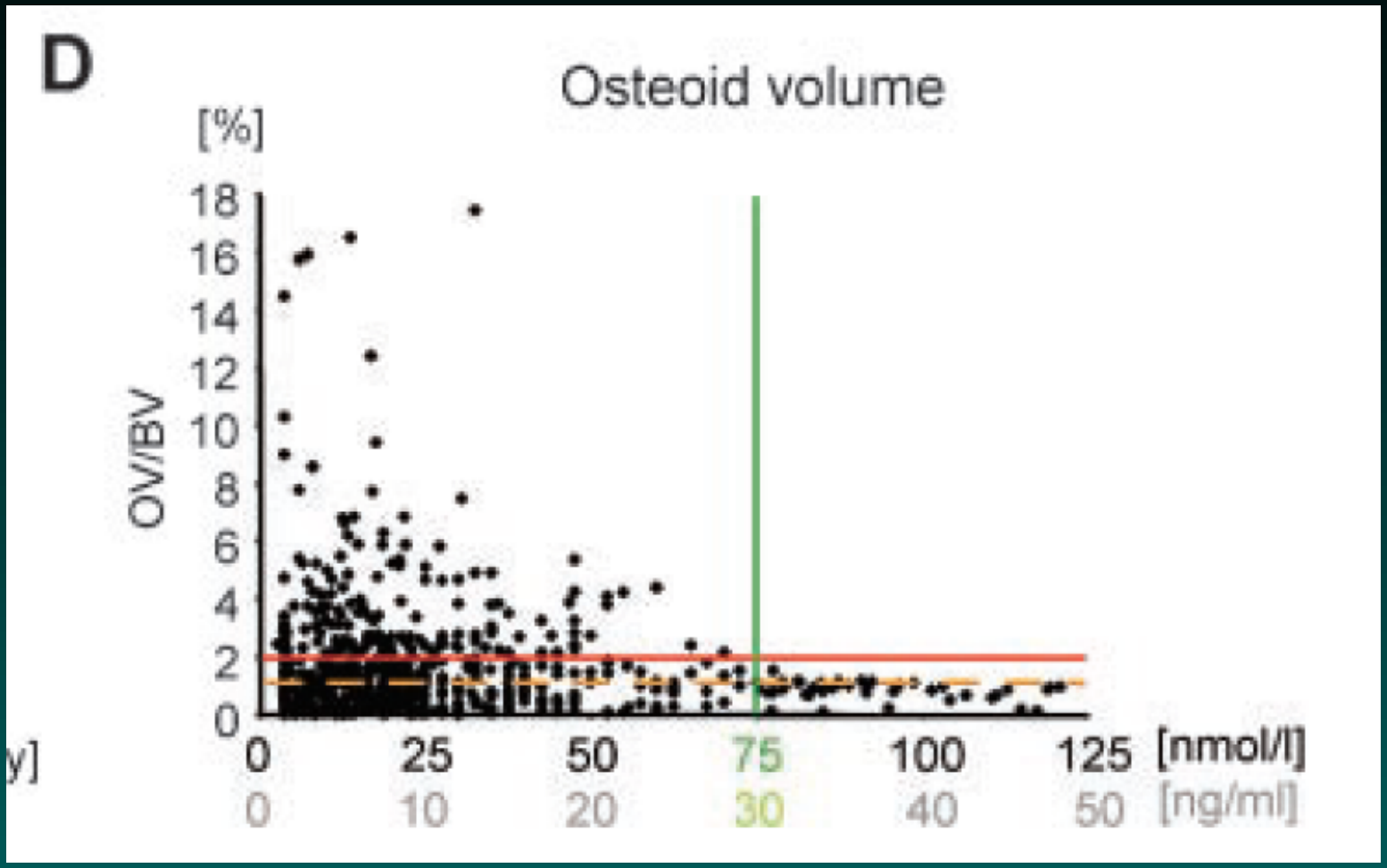
$$P(req\,met|intake) = \int_{serum} P(req\,met|serum) * P(serum|intake)$$

What is P(Req Met | Intake)?

Let Y be the serum level attained at a given intake.
Let X be the serum level required.
Assuming (uncorrelated!) bivariate normality,

$$
\begin{aligned}
P(Y > X) &= P\left(\frac{Y-\mu_Y}{\sigma_Y} > \frac{X-\mu_Y}{\sigma_Y}\right) \\
&= P\left(Z_Y > \frac{(X-\mu_X)-(\mu_Y-\mu_X)}{\sigma_X * (\sigma_Y/\sigma_X)}\right) \\
&= P\left(Z_Y - (\sigma_X/\sigma_Y) * Z_X > \frac{\mu_X-\mu_Y}{\sigma_Y}\right) \\
&= P\left(Z > \frac{\mu_X-\mu_Y}{\sqrt{\sigma_Y^2+\sigma_X^2}}\right)
\end{aligned}
$$

Priemel et al, Figure 4d: OV/BV



OV/BV values $\geq$ 1.2% or 2% are bad.

Priemel et al recommended targeting $>$30 ng/mL.

The IOM Committee's Cutpoint

20 ng/mL = 50 nmol/L.

Why? Because 97.5% isn't 100%.

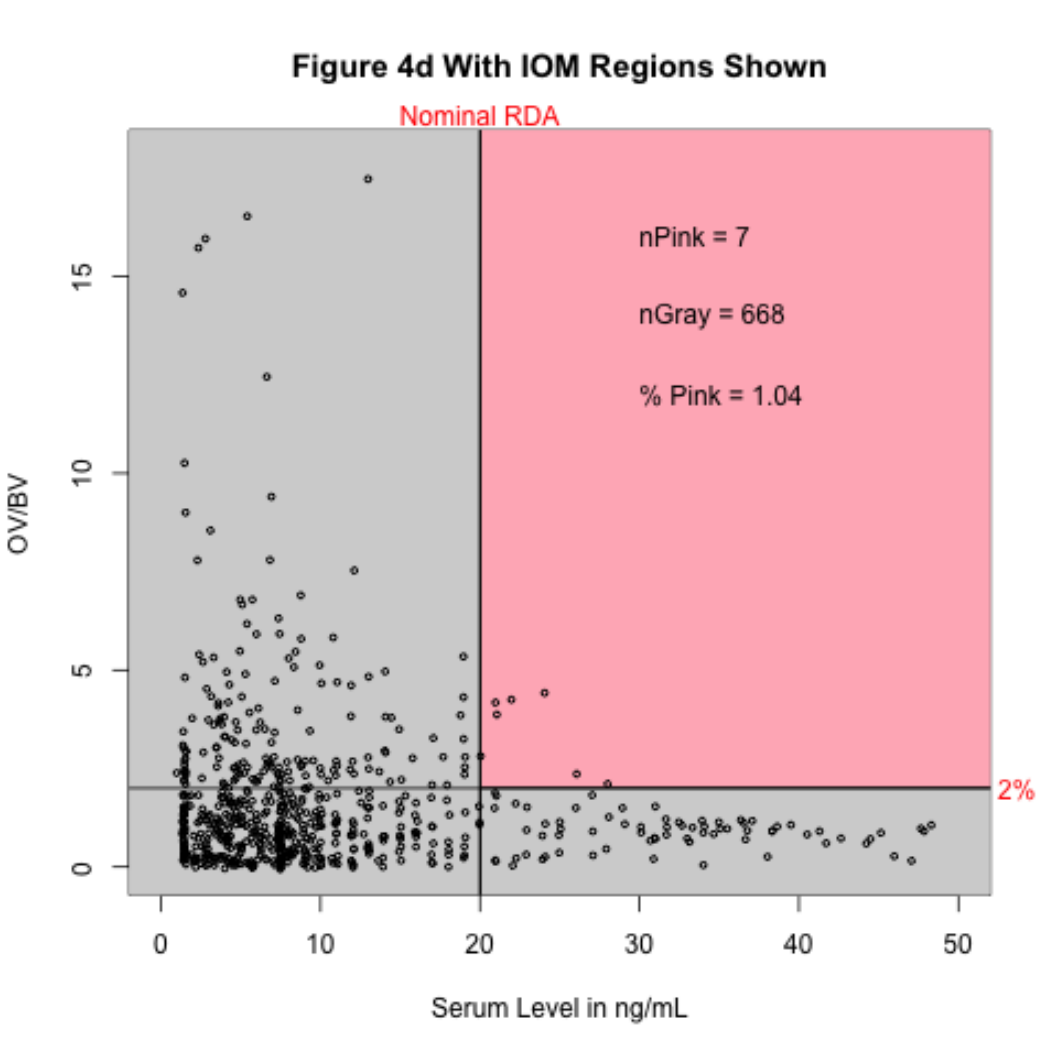
"The number ... above 50 nmol/L was counted by inspection...

At ... 50 nmol/L, there were seven data points reflecting ... (OV/BV $>$ 2 percent).

This suggested ...
50 nmol/L met the needs of 99 percent ... (that is, only 7 of 675 surpassed the measure)."

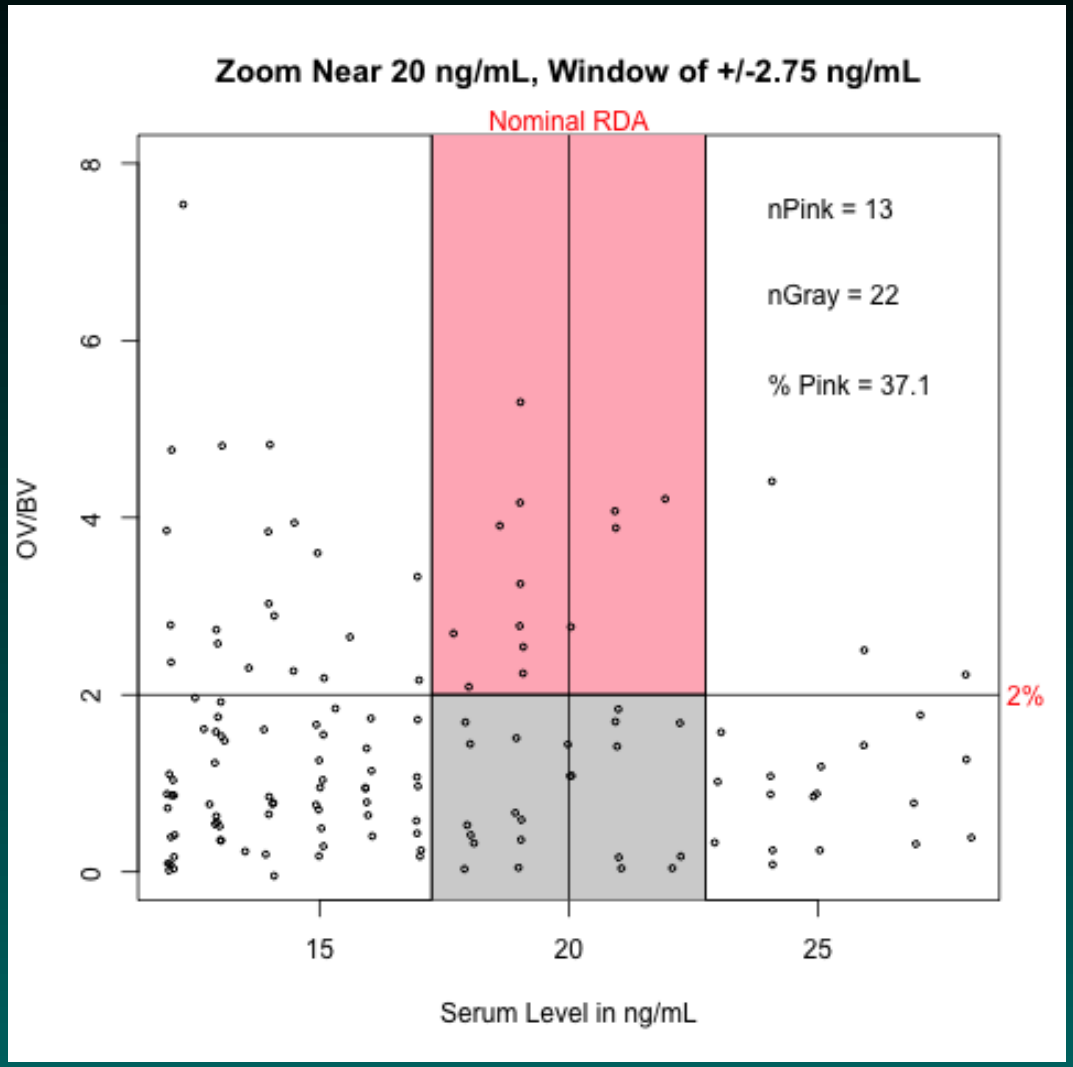
IOM report, p.276.

Come Again?



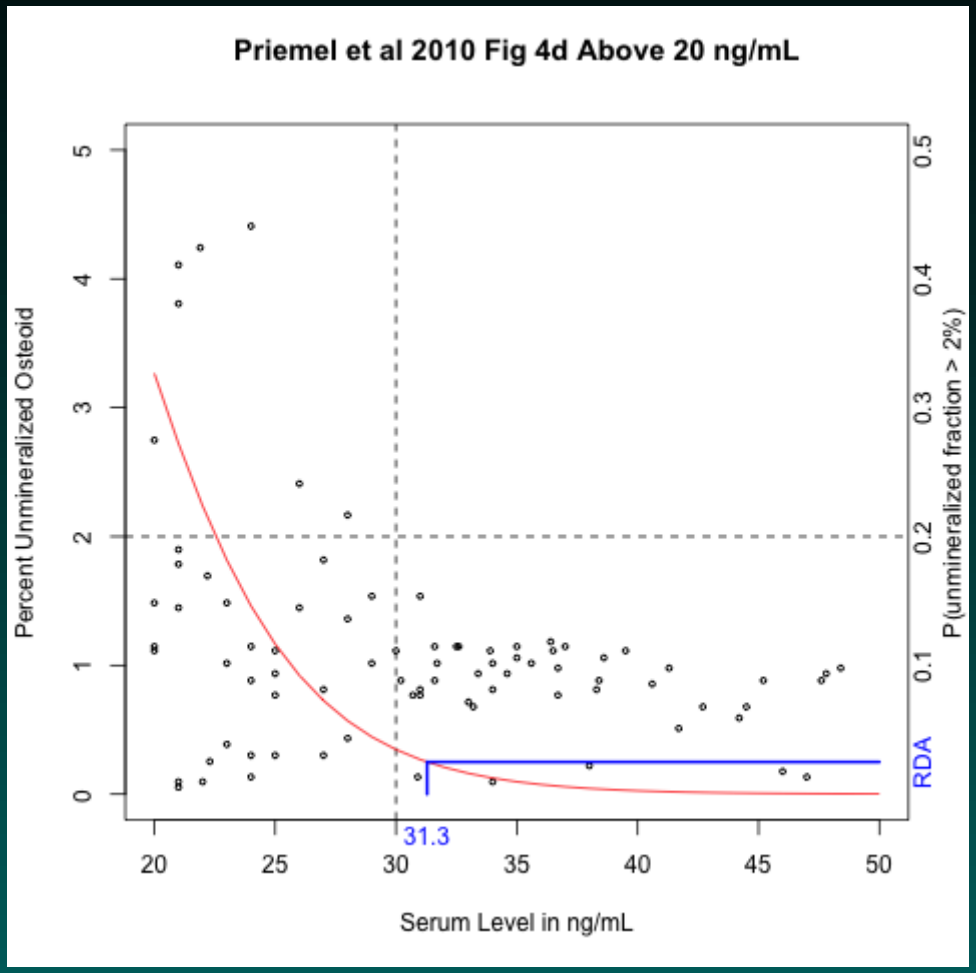
Is this picture reasonable?

Zooming In on 20 ng/mL



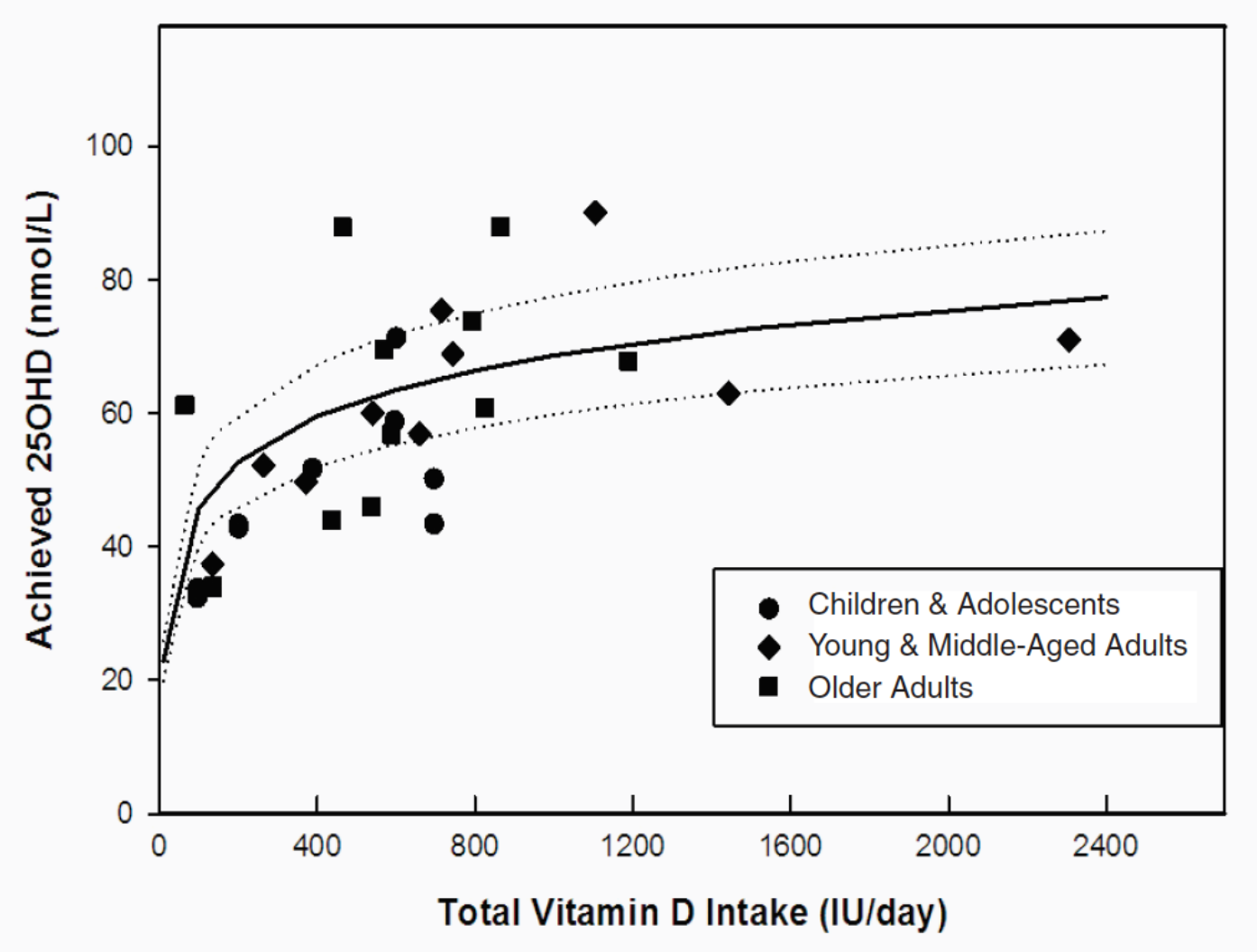
This rate of problems is way too high

What if we Fit a Model?



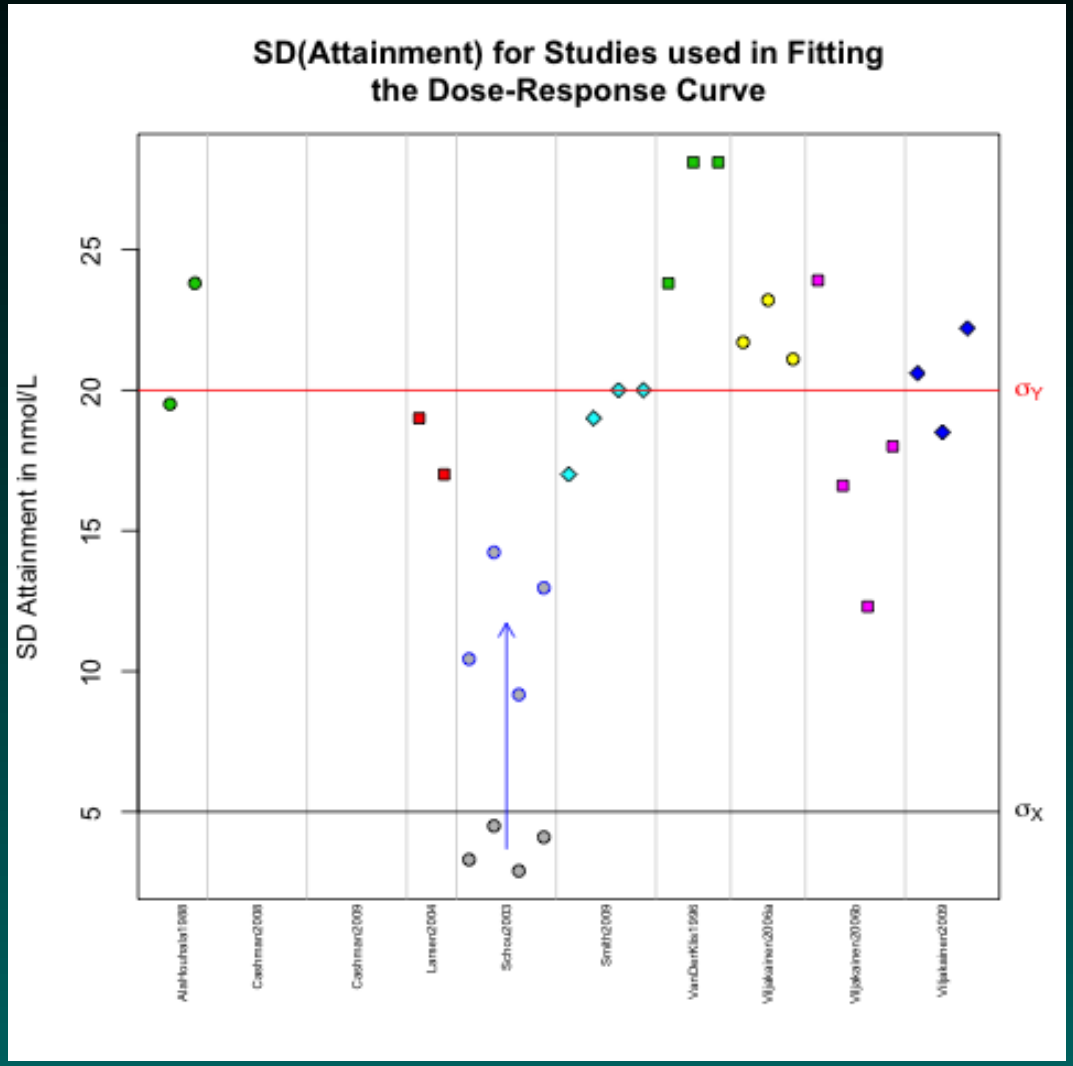
Logistic regression suggests a cutpoint nearer to 30.
What about attainments?

Mapping Serum to Intake, IOM 2011 Fig 5.4



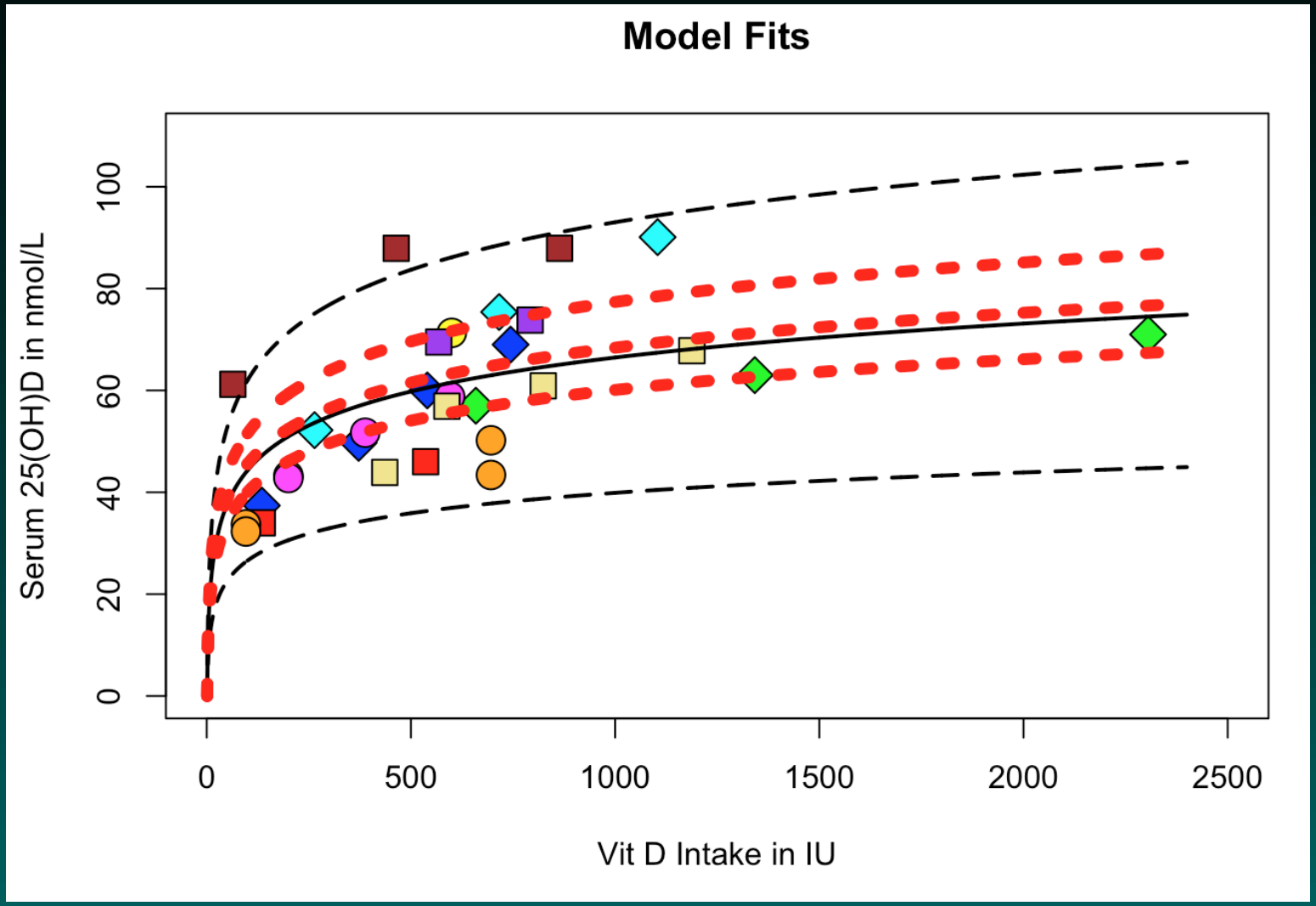
Are cohort averages (dots) vertically close to a curve?

SD(Attainments) for the Studies Used



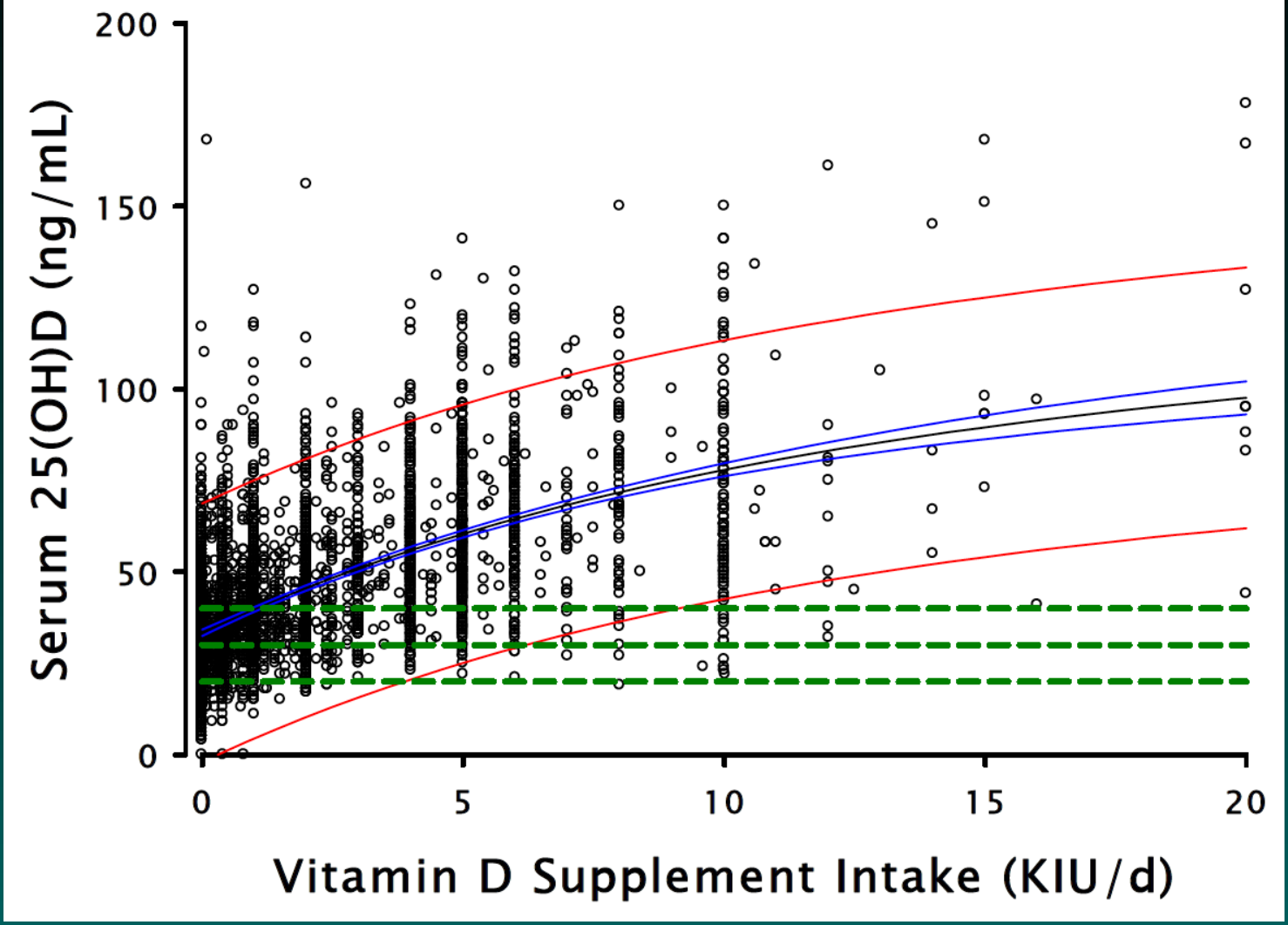
Here, $\sigma_Y/\sigma_X \approx 4$. (One study used SEM.)

All the Data



IOM in red, About 2 SEM; prediction in black.

We Have Observational Data for Serum | Intake



Heaney et al Figure 1; 3,657 participants (GrassrootsHealth)

We have Theory for Req Met | Serum

Assuming

$P(reqs\ met|serum)$ is $N(23, 2.5)$

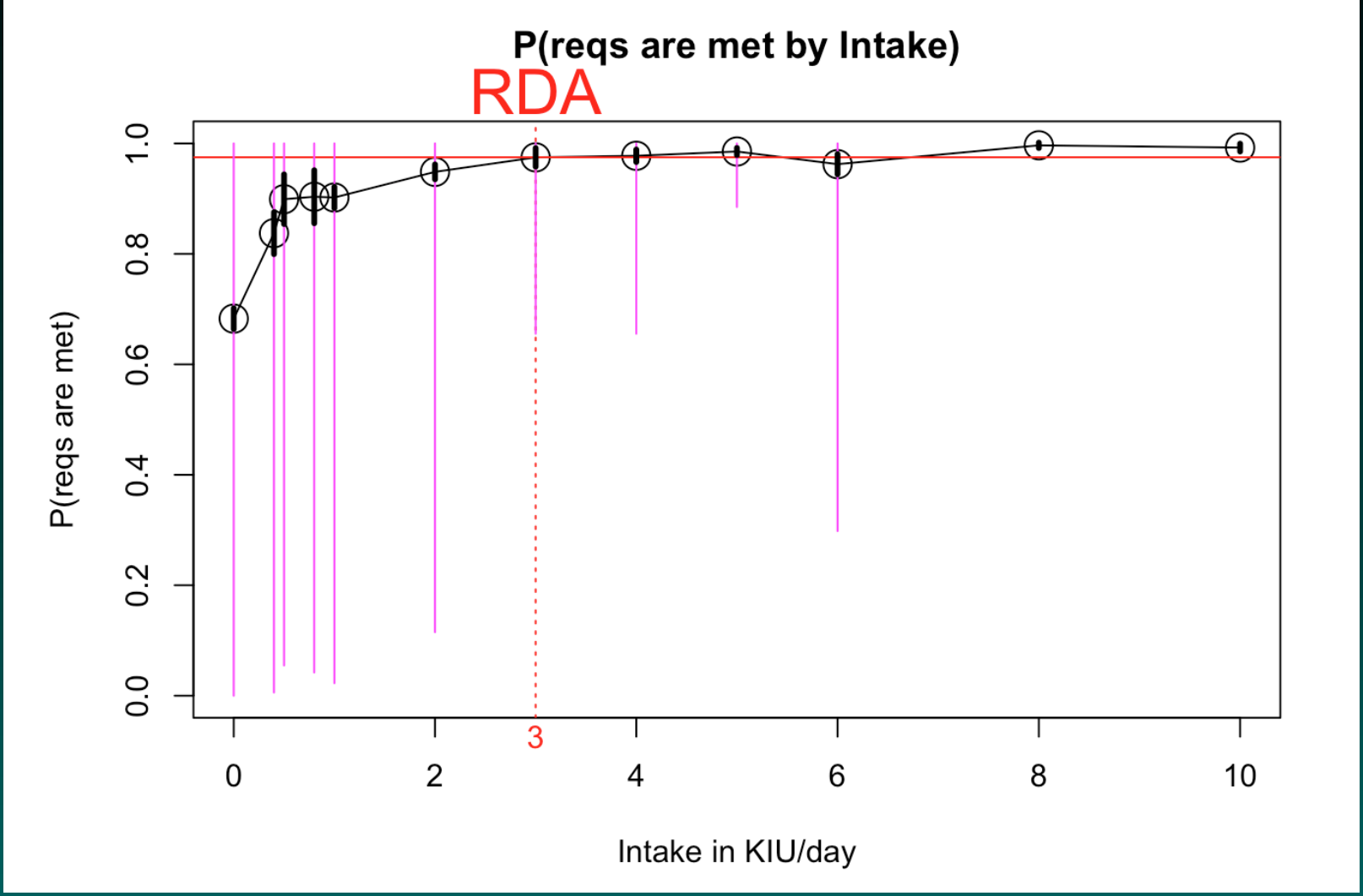
based on iPTH results (Joseph et al, 2016, unpublished), and using

empirical distributions of $P(serum|intake)$

at every intake GRH measured ≥ 100 times, we can

estimate the intake where $P(reqs\ met|intake) \geq 0.975$ (the RDA).

Our Estimate



We get about 3K IU/day.

Where We Are Today (1/2)

It's now 6 years since these recommendations went into effect.

Things have been rather contentious in the interim.

We raised these arguments with the Report Review Committee (RRC) of the National Academies of Science, Engineering and Medicine (NASEM) late last year.

I gave a presentation on this at an NASEM meeting on Reproducible Research in March

Where We Are Today (2/2)

The NASEM appointed an expert panel to review our claims in March. This panel reported back to NASEM leadership at the end of May.

The report is not yet publicly available.

Where We Are Today (2/2)

The NASEM appointed an expert panel to review our claims in March. This panel reported back to NASEM leadership at the end of May.

The report is not yet publicly available.

The NASEM has now appointed a second panel to review what the RDA should be in light of the first panel's findings.

The second panel's report is expected in August. NASEM leadership will comment publicly shortly afterwards.

The math isn't really the hard part here. Clearly documenting what was actually done is.

Acknowledgments

Porter Coggeshall, Alicia Carriquiry

GrassrootsHealth

Carole Baggerly, Sharon McDonnell

Carol Wagner, Bruce Hollis (MUSC)

Robert Heaney (in memoriam)

Stephen Feinberg (in memoriam)