

Lies, Damned Lies,

And What To Do With Them

Ken Rice, CHRU Work in Progress

March 2, 2005

A question for English majors

“There are three types of lies;

lies, damned lies, and statistics”

Which famous philistine said this?

A) Benjamin Disraeli
B) Mark Twain
C) George Bernard Shaw
D) Oscar Wilde

A question for English majors

"There are three types of lies;

lies, damned lies, and statistics"

Which famous philistine said this?

A) Benjamin Disraeli
B) Mark Twain
C) George Bernard Shaw
D) Oscar Wilde

E) Any of them, depending on who you ask (a Bayesian answer)

Structure of the talk

- Lies (from you)
- Damned lies (from data)
- Statistics (from me)

Raise your hand... if appropriate

Q.

Raise your hand... if appropriate

Q. Are you having an extra-marital affair?

Raise your hand... if appropriate

Q. Are you having an extra-marital affair?

Q. Have you taken money from drug companies?

Raise your hand... if appropriate

Q. Are you having an extra-marital affair?

Q. Have you taken money from drug companies?

Q. Have you had an STD/driven drunk/taken drugs...

A. We are unlikely to get 100% truthful answers

Can't always collect 'perfect' data – but using a **statistical model** can help. (Get a coin ready)

A fun game about lying



1. Toss 2 coins – keep results **secret**

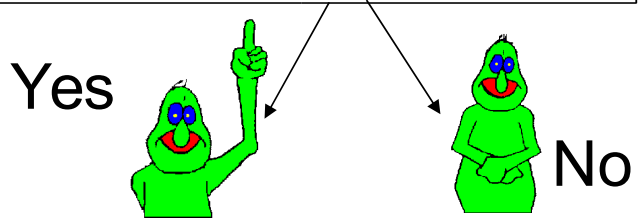
2 Heads

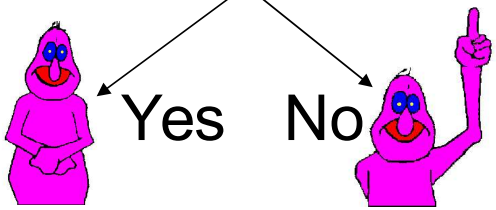
0,1 Heads

2. Rude Q:

Tell the truth!

Lie!





3. Math: $p/4$ $(1-p)/4$ $3(1-p)/4$ $3p/4$

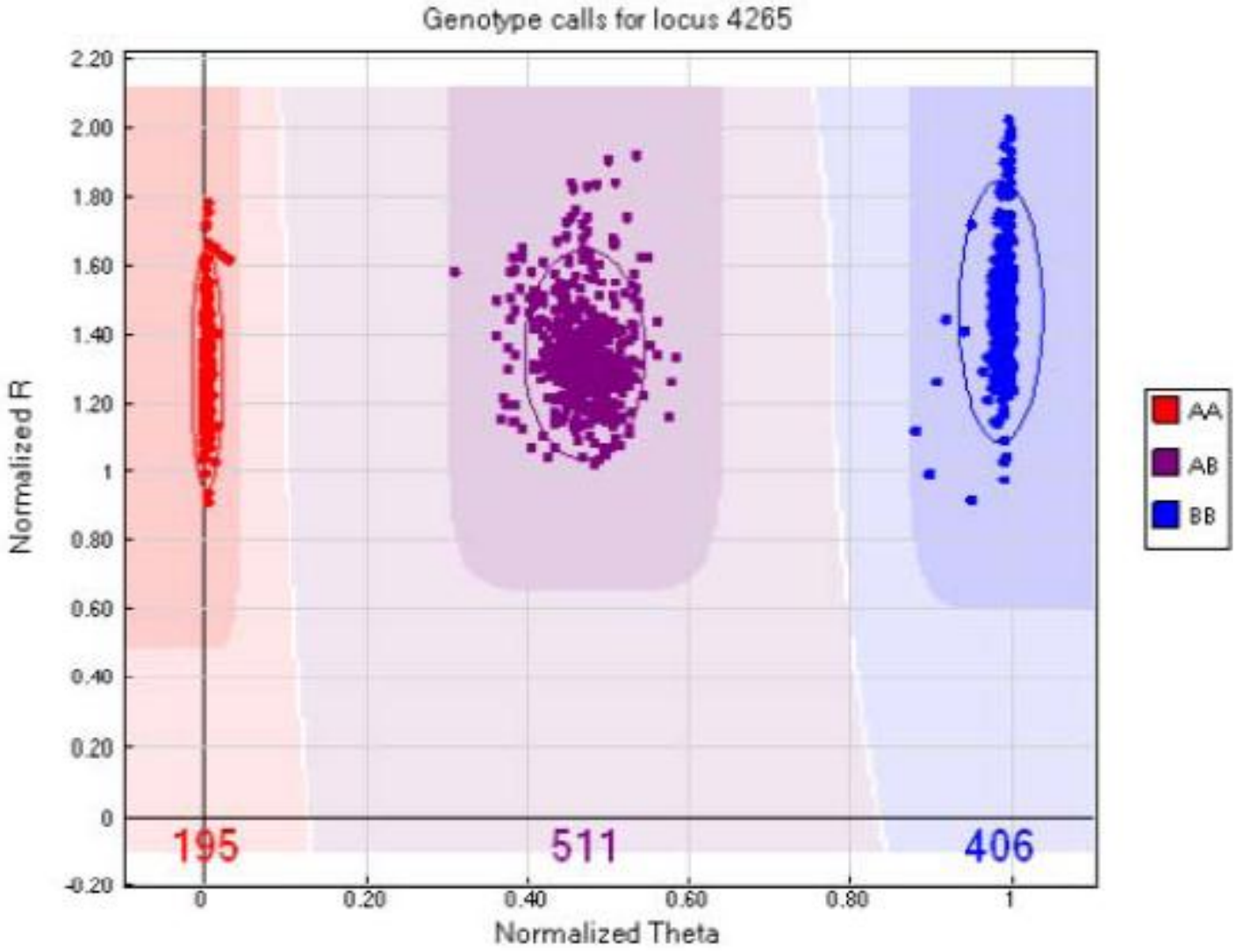
Proportion of hands up is about $3/4 - p/2$

What did we learn from that?

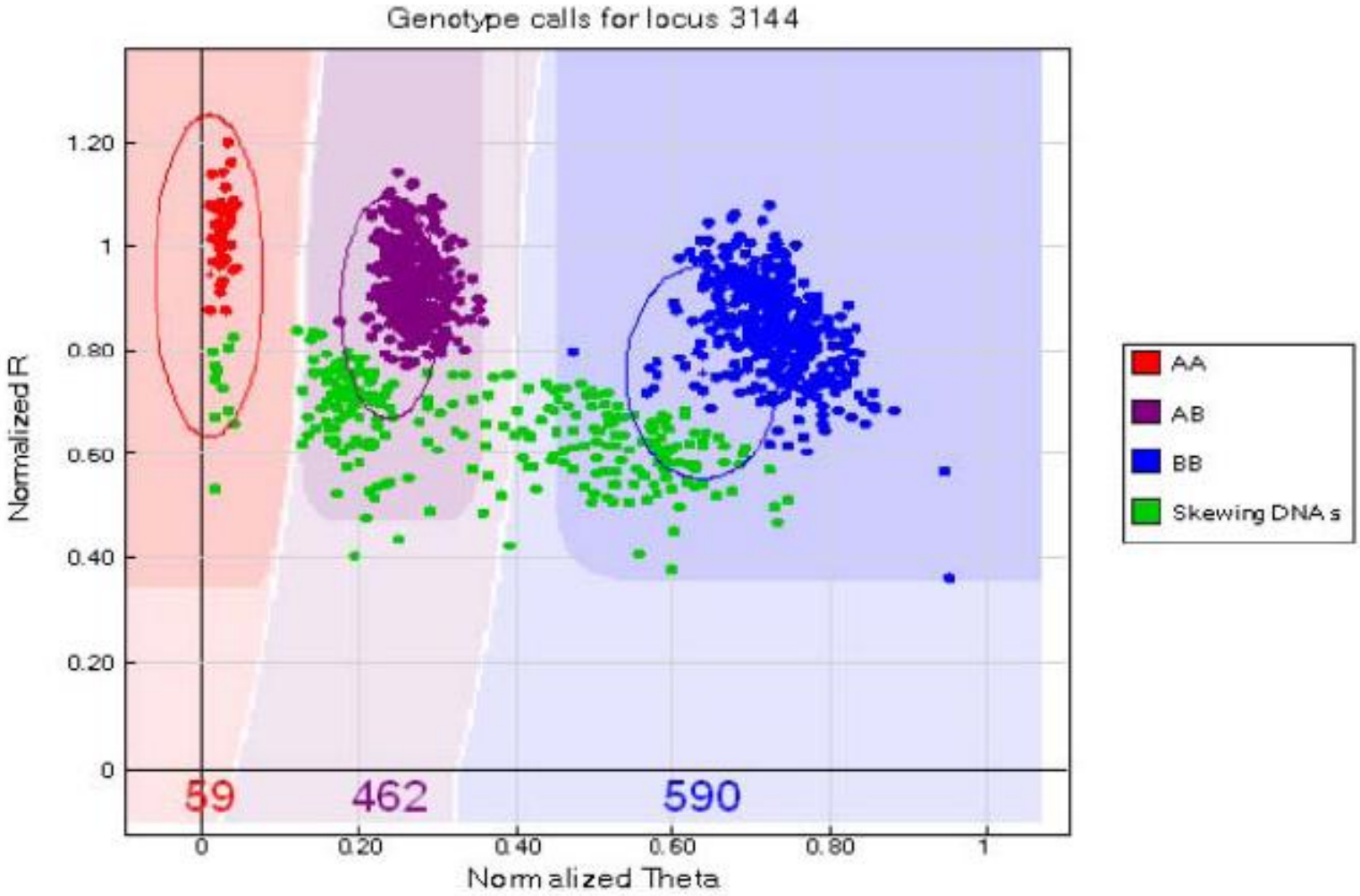
This is a 'randomized response' technique (Warner 1965). We get a **misclassified** version of the truth – and by design we can't uncover the 'truth' for each person.

- The model for **misclassification** is important;
  - The 'truth' actually means something
  - Your coins are all 'fair' - 50:50 heads/tails
  - You followed the flowchart
  - You weren't influenced by anyone else
- Have to make **assumptions** like this to get much done (e.g. estimate p)
- If you know **nothing** about 'fairness' you learn **nothing** about p
- Typical 'solutions' are more assumption-driven than we'd usually like

Some very nice genotype data



Some damned lies!



Ignoring the problem, or the whole dataset, is **A Bad Idea**

Assumptions about the error process

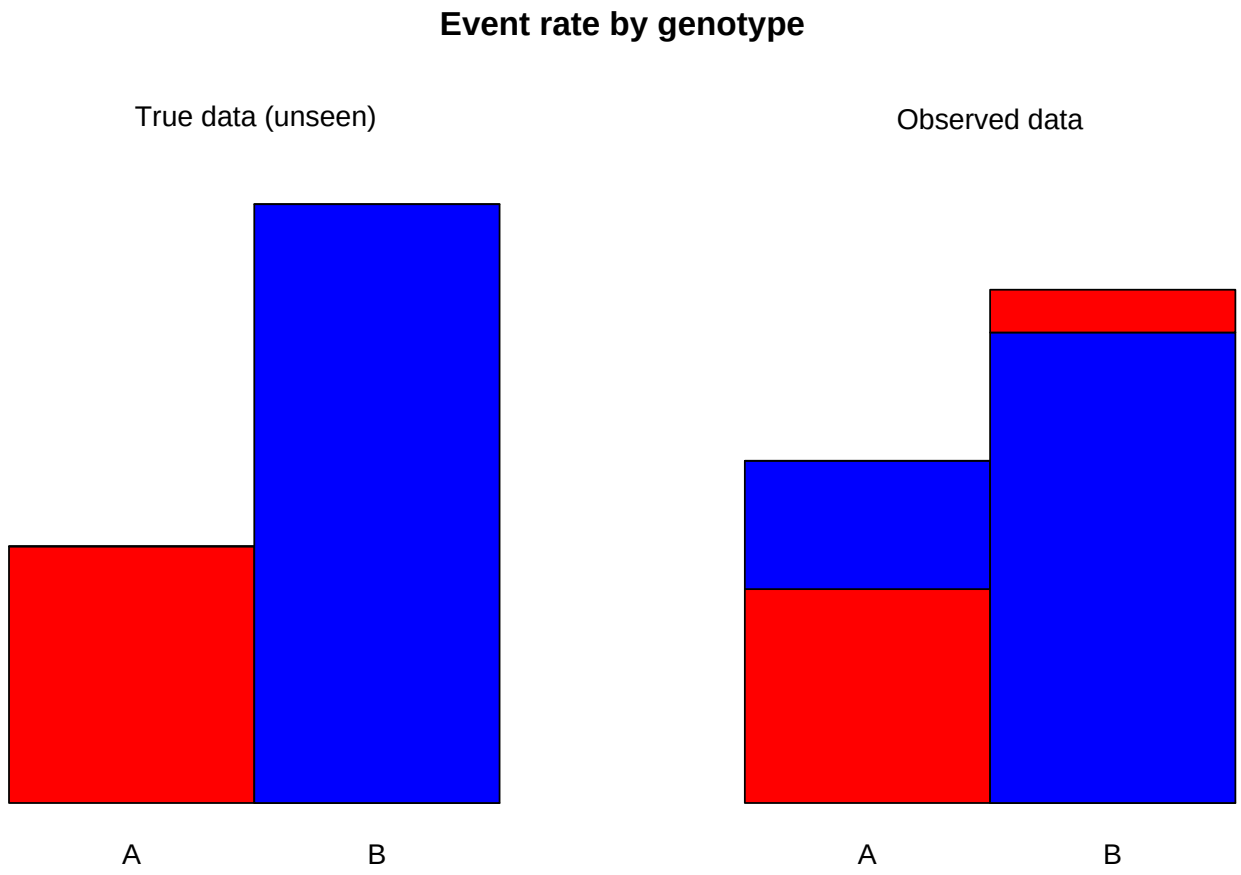
These are basically the same as in the coin-tossing game;

- Everyone has a ‘true’ genotype
- Only your **true** genotype affects your **observed** genotype (if anything)
- Process is same for cases/controls, men/women etc
- So-called **non-differential** misclassification
- Realistic for genotyping error, less so elsewhere (drinkers/smokers)

What practical effect will the error have?

Bias towards the null

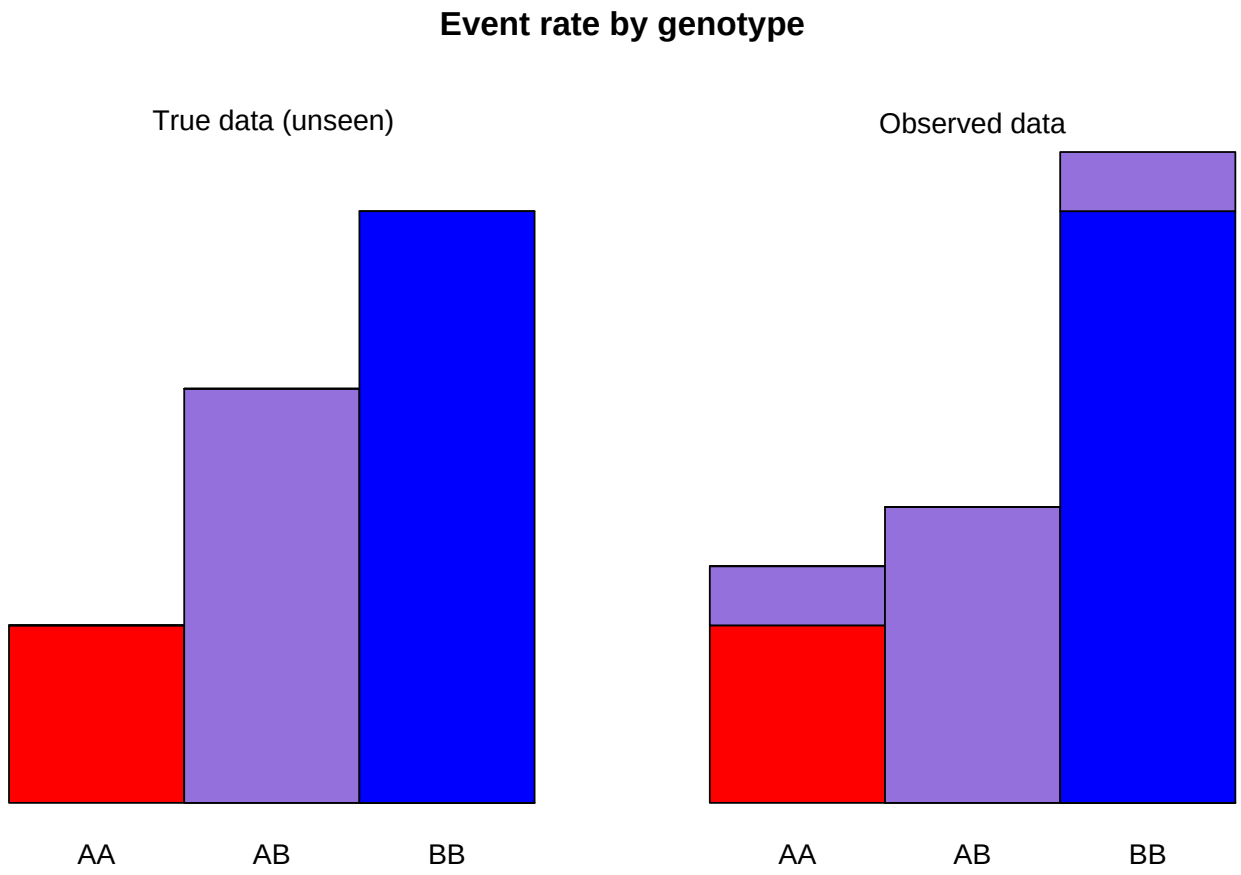
For a binary exposure, non-differential misclassification biases **towards the null**



Should **expect** estimates to get ‘more extreme’ - intervals less intuitive

Bias in lots of directions

For **any** other situation intuition is elusive;



Claiming ‘bias towards the null’ here is unhelpful and often inaccurate

Unmatched case-control setup

Switch to a retrospective analysis;

	Observed Genotype		
Disease Status	AA	AB	BB
Control	p_1	p_2	p_3
Case	q_1	q_2	q_3

- Perhaps want to **test** equality, where $p = q$
- Or **estimate** odds ratios; $\frac{q_2 p_1}{q_1 p_2}$ and $\frac{q_3 p_1}{q_1 p_3}$

Normally the two approaches would be extremely similar – consider them both under misclassification;

Testing equality ($p = q$) under misclassification

- Assume non-differential misclassification, known error rates (sensitivity, specificity)
- **Likelihood Ratio** tests are **totally** unaffected (Bross, 1954)

Testing equality ($p = q$) under misclassification

- Assume non-differential misclassification, known error rates (sensitivity, specificity)
- **Likelihood Ratio** tests are **totally** unaffected (Bross, 1954)
- Not as wacky as it seems (Rice & Holmans, 2003) - still just testing $p = q$
- Exact/mediocre/zero information on error rates... same result!
- Not a panacea: Type I error stays the same but **power** is reduced
- Doesn't hold if we impose structure on p, q

Estimation (regression) under misclassification

- Misclassification alters the (imposed) 'structure' of p, q
- 'How much' depends on the error **rates** θ (sensitivity, specificity)
- We modify our 'basic' model appropriately (not trivial);

	Observed Genotype		
Disease Status	AA	AB	BB
Control	$\theta_{11}p_1 + \theta_{12}p_2 + \theta_{13}p_3$	...	...
Case	$\theta_{11}q_1 + \theta_{12}q_2 + \theta_{13}q_3$	...	...

- Apply your favorite statistical **method** (MLE/GEE/Bayes)
- With good knowledge of error rates θ, and 'well-behaved' data, estimation is not too contentious

Badly behaved data

	Genotype AA	Genotype AB	Misclassification?	Estimate
Observed data	10	90	Ignore it!	$p_{AA} = 10\%$
Possible 'truth' (nice)	5	95	Sens, Spec $\approx 95\%$	$p_{AA} = 5\%$
Possible 'truth' (not nice)	0	100	Sens, Spec $\approx 90\%$	$p_{AA} = 0\%$

- If the data 'point to' a zero in p, q, odds ratios go to zero, infinity
- Report this! – awkward, but 'fixes' are worse
  – unless you're **really** Bayesian (Rice, 2003)
- For finite samples, rare genotypes, **not** unrealistic behavior

Confidence intervals for estimates

- Probably of more interest than estimate!
- Again, well-behaved data allow use of standard **methods** (though not standard software)
- Most methods 'don't work' when OR estimate is infinite
- 'Naive' interval coverage is (very roughly) 95% $\times$ $\mathbb{P}$(data was nice)
- For infinite estimates, the data at best **really** just give an upper/lower bound. Report this! (not a fudge)

Really **badly behaved data**

- Odds **ratios** can go to zero, infinity
- They **can** also go to 0/0 (even for a 2x2 table)
- This is **undefined!**
- No estimate or interval makes **any** sense
- The data are telling you they can't answer your question **at all**

Different models for regression

Plain regression



Under misclassification



Different models for regression

Plain regression



Fast
Generic, Reliable
Easy to use
Same power as everyone else

Under misclassification



Slow
Problem-specific, Can behave oddly
Requires expertise (perhaps extra data)
Extra power available

Misclassification can make us think harder about the output, but we still just fit models

Regression Methods which *aren't* really models

Some 'non-standard' techniques don't fit 'full' models (e.g. Conditional Logistic Regression, Cox Regression)

In vehicular form;

'Non-standard' regressions

Regression Methods which *aren't* really models

Some 'non-standard' techniques don't fit 'full' models (e.g. Conditional Logistic Regression, Cox Regression)

In vehicular form;

'Non-standard' regressions



- You can still 'go places' but in a rather different way.
- Limited comparisons with full models – perhaps fast/slow?
- **Very** hard to extend for misclassified data
- Car → Truck Not **too** hard
  Sub → Truck ???

We really need one of these!





Finding ‘full model’ ways of explaining non-standard methods is useful (but hard);

- Helps explain *why* non-standard methods work so well
- Much easier to allow for misclassification;
- Sub → Truck is just Car → Truck
- Just *really* cool

Examples of this re-interpretation

- Matched case-control studies, binary exposure (Rice, 2004)
- Pair-matched case-control studies, multi-category exposure (submitted)
- Genetic Trio studies (Clayton & Cordell method) (in progress)
- Cox regression (in progress)

Enables adjustment for misclassification (e.g. Rice, 2003) but also missing data – and maybe more?

Coronary Calcium Scoring



- White ‘blobs’ in heart
- Agatston scoring method
- This is **far** from perfect
- Most analysis currently ignores this;
  - ‘Agatston $>$ 0’ $\sim$ genotype
  - CHD $\sim$ Agatston
- How much does it matter?

Uncertainty in Agatston score

Just how bad is it? Scan each person twice;

		Agatson Score 1				
		<0.5	0.5-10	10-100	100-400	400+
Score 2	<0.5	3388	103	42	2	0
	0.5-10	97	190	85	0	1
	10-100	34	87	1056	88	0
	100-400	0	0	86	773	39
	400+	0	0	0	41	632

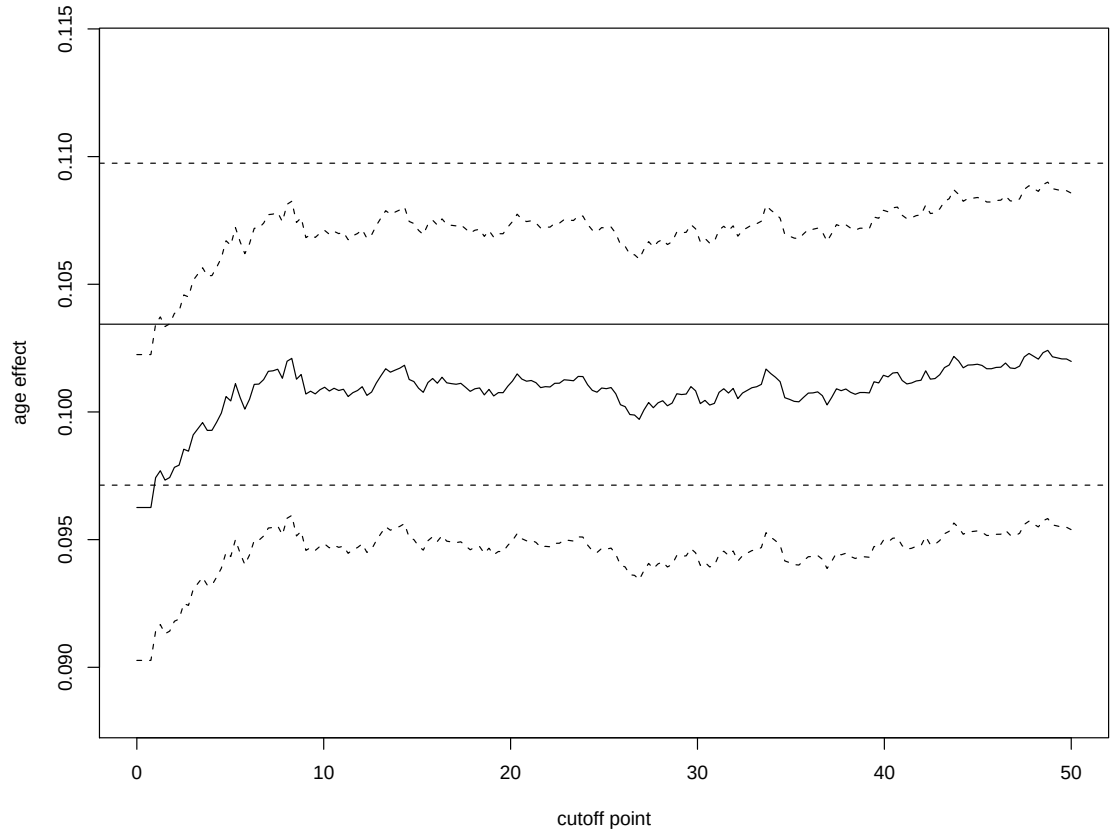
- Clearly **some** unreliability
- Broad agreement with other studies

Agatston score

- Suppose we care about Agatston =0, $>$0 only
  – simple approach takes **average** score
- Perhaps **both** scores are misclassified version of unseen ‘truth’?
- Can estimating effects of e.g. age, gender, blood pressure...
- Get error rates ‘for free’ (typically 96-99%)
- **But**
  - ‘Non-differential’ probably not right (in progress)
  - ‘True’ Agatston score a slippery concept
- Imperfect, but a reasonable sensitivity analysis

Agatston score cutoff

We used Agatston =0, $>$0. What about Agatston $<$10, $>$10? 20? 50?



Misclass-adjusted estimate a bit higher, same ball-park **change** in estimates – never a huge effect

In summary

- There **are** lies (and damned lies!)
- Also statistics
- The latter can help interpret the former

“It is easy to lie with statistics.
It is hard to tell the truth without it.”

Andrejs Dunkels

References

- Warner SL (1965) A survey technique for eliminating evasive answer bias, JASA 60:63-69
- Dosemeci M, Wacholder S, Lubin JH (1990) Does nondifferential misclassification always bias a true effect toward the null value? Am J Epidemiol 132:746-748
- Bross L (1954) Misclassification of 2x2 tables Biometrics 10:478-486
- Rice K (2004) Equivalence between conditional and mixture approaches to the Rasch model and matched case-control studies, with applications. JASA, 99(466):510-522
- Rice K (2003) Full-likelihood approaches to misclassification of exposure in matched case-control studies. Stats in Med, 22(20):3177-3194
- Cordell H, Barratt B, Clayton DG (2004) Case/pseudocontrol analysis in genetic association studies. Gen Epi, 26(3):167-185
- Rice K, Holmans P (2003) Allowing for genotyping error in analysis of unmatched case-control studies. Annals of Human Genetics, 67(2):165-174
- `http://faculty.washington.edu/kenrice`