BIOST 111

Statistical Consulting:

Where it helps, and where it might have

Ken Rice

UW Biostatistics

June 3, 2009

Some Research Interests

Like most faculty, my time is split various ways;

- Cardiovascular disease; genetic variations, other risk factors
- Multiple comparisons, outlier-detection
- Measurement error problems
- 'Translating' frequentist $\Longleftrightarrow$ Bayesian methods

Also teaching! 1st year course in R, 2nd year 'advanced regression'. The following examples represent our **consulting** course, taken in 2nd/3rd year.

Some Research Interests

Consulting and related areas;

- Some cool applications we *did* help with
- Some applications where statistical consulting *would* have helped
- The unfriendly world of academic publication

Consulting: A mountain to climb

Ray Huey (UW Biology) has fantastic records on climbers attempting Mt Everest – summiting or not, also dying on the ascent/descent

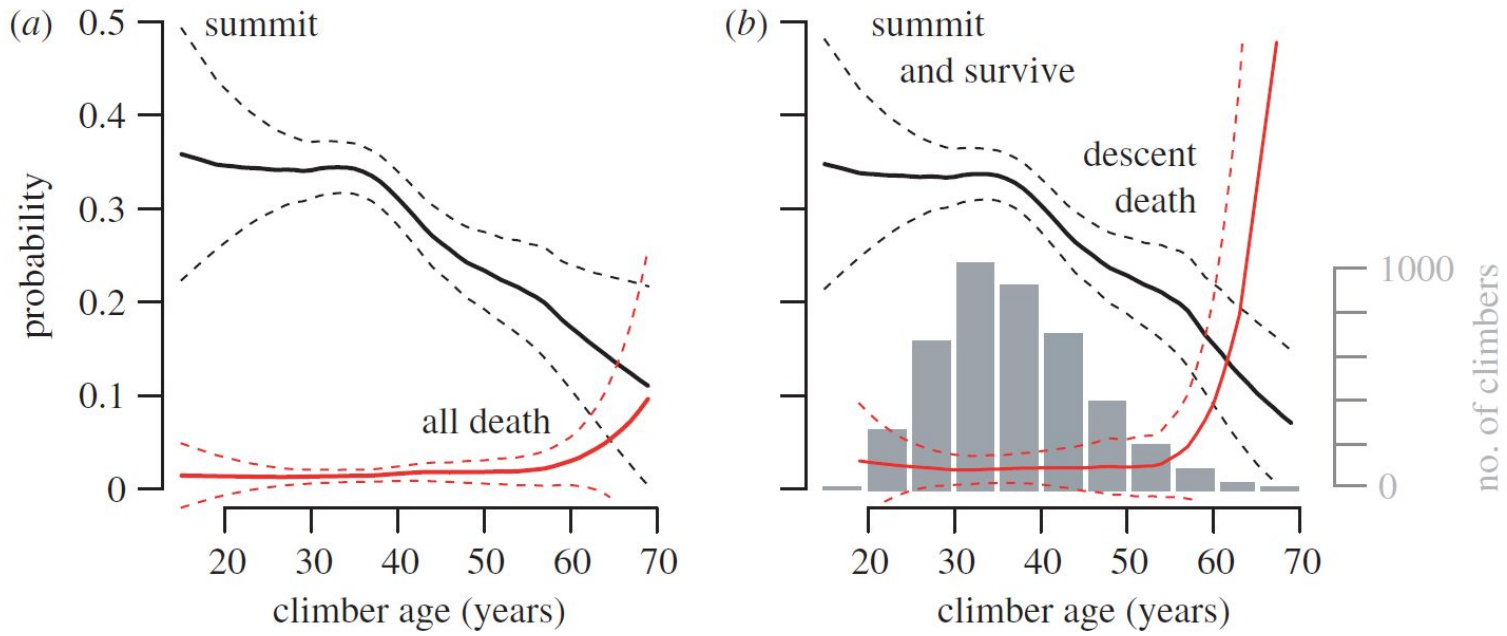
- 20+ years of data; age, route, nationality, sex
- Can these risk factors be separated?

Consulting: A mountain to climb

biology letters
Evolutionary biology

Biol. Lett. (2007) **3**, 498–500
doi:10.1098/rsbl.2007.0317
Published online 14 Augsut 2007

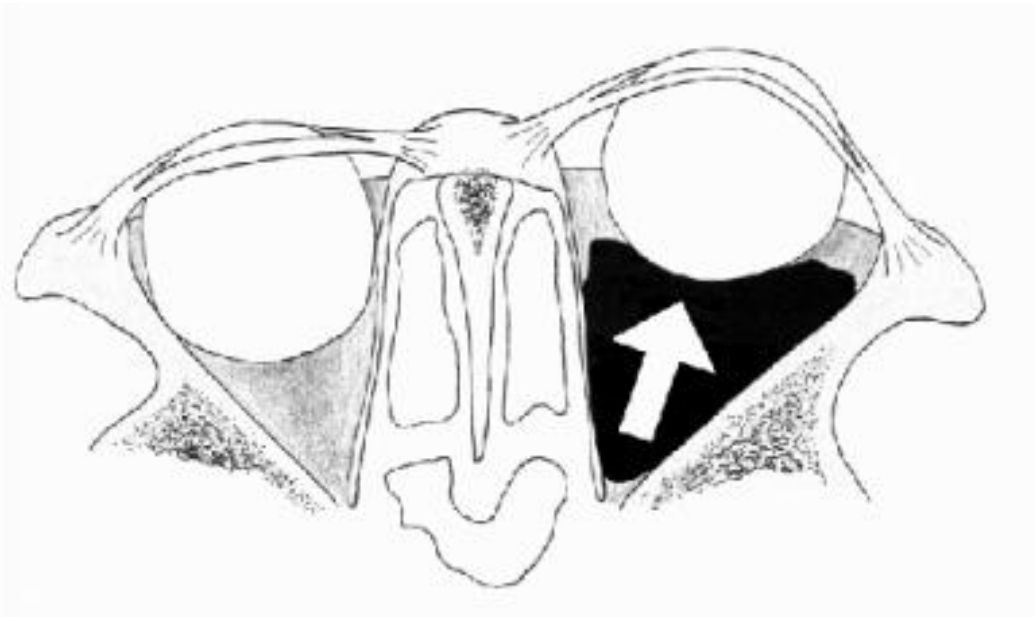
Effects of age and gender on success and death of mountaineers on Mount Everest

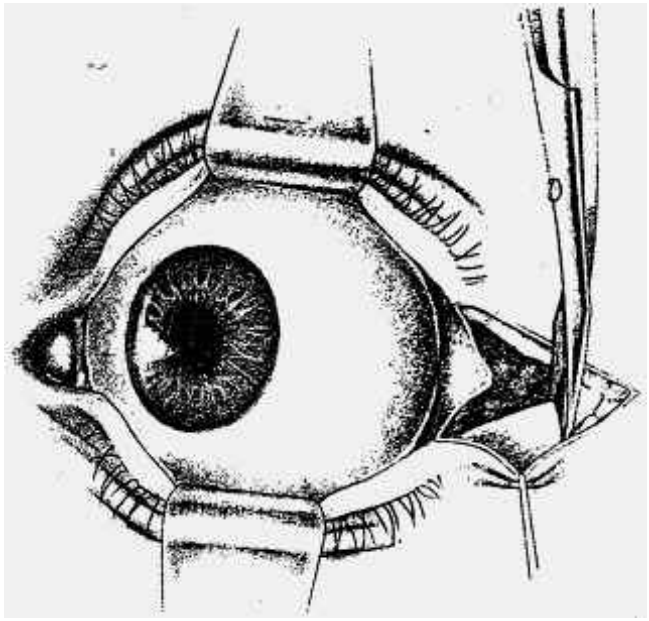


- Gender looks unimportant, old age is risky
- ...particularly on the way down
- Experience helps you summit (not live)

Consulting: Eyeball the data

Burns patients receive a lot of fluids; do they get too much?

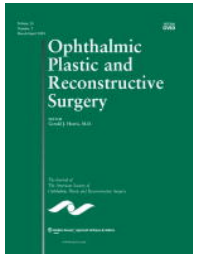




These are the least-gross pics available!

- *Does* it harm? a.k.a. 'Orbital Compartment Syndrome'
- Patients have two eyes but only one IV drip; n=28 or n=56?
- Are we getting the most from this data?

Consulting: Eyeball the data



Orbital Compartment Syndrome in Burn Patients

Christopher N. Singh, M.D.*, Matthew B. Klein, M.D.†, Stephen R. Sullivan, M.D.†, Bryan S. Sires, M.D., Ph.D.*, Carolyn M. Hutter, M.S.‡§, Kenneth Rice, Ph.D.‡, and Arash Jian-Amadi, M.D.*

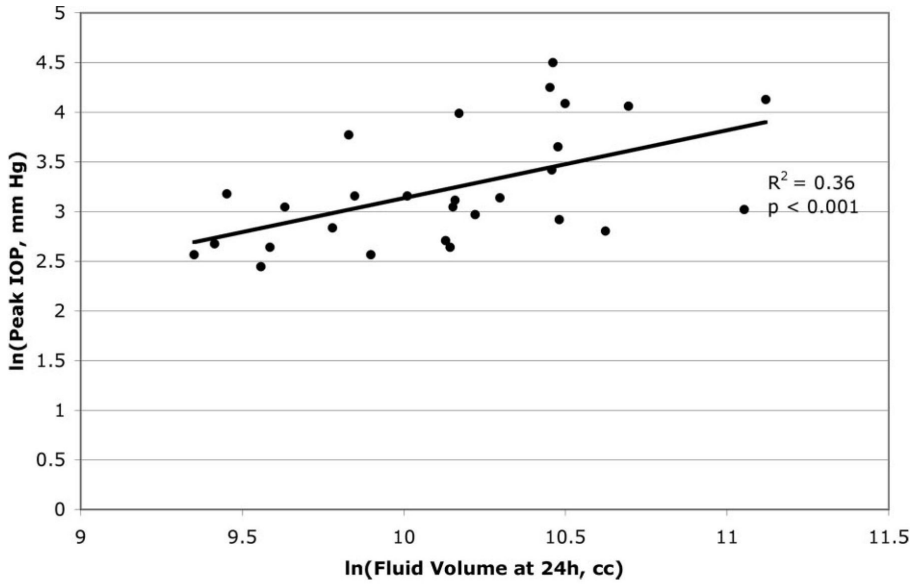


FIG. 1. Peak IOP (mm Hg, natural log) at 48 hours compared with fluid volume administered at 24 hours (ml, natural log) for all patients (n = 28).

- Effectively, n=28 (certainly not *much* more)
- Increased pressure, yes; damage is not so clear-cut
- This *is* about all the data can tell us – which is re-assuring

Caution: Bee careful

What happens when you introduce hybrid bees into new environments? (Roubik, 1978)

Competitive Interactions Between Neotropical Pollinators and Africanized Honey Bees

Abstract. *The Africanized honey bee, a hybrid of European and African honey bees, is thought to displace native pollinators. After experimental introduction of Africanized honey bee hives near flowers, stingless bees became less abundant or harvested less resource as visitation by Africanized honey bees increased. Shifts in resource use caused by colonizing Africanized honey bees may lead to population decline of Neotropical pollinators.*

SCIENCE, VOL 201, 15 SEPTEMBER 1978

(There was a lot of *buzz* about this paper...)

Caution: Bee careful

Here's the big result;

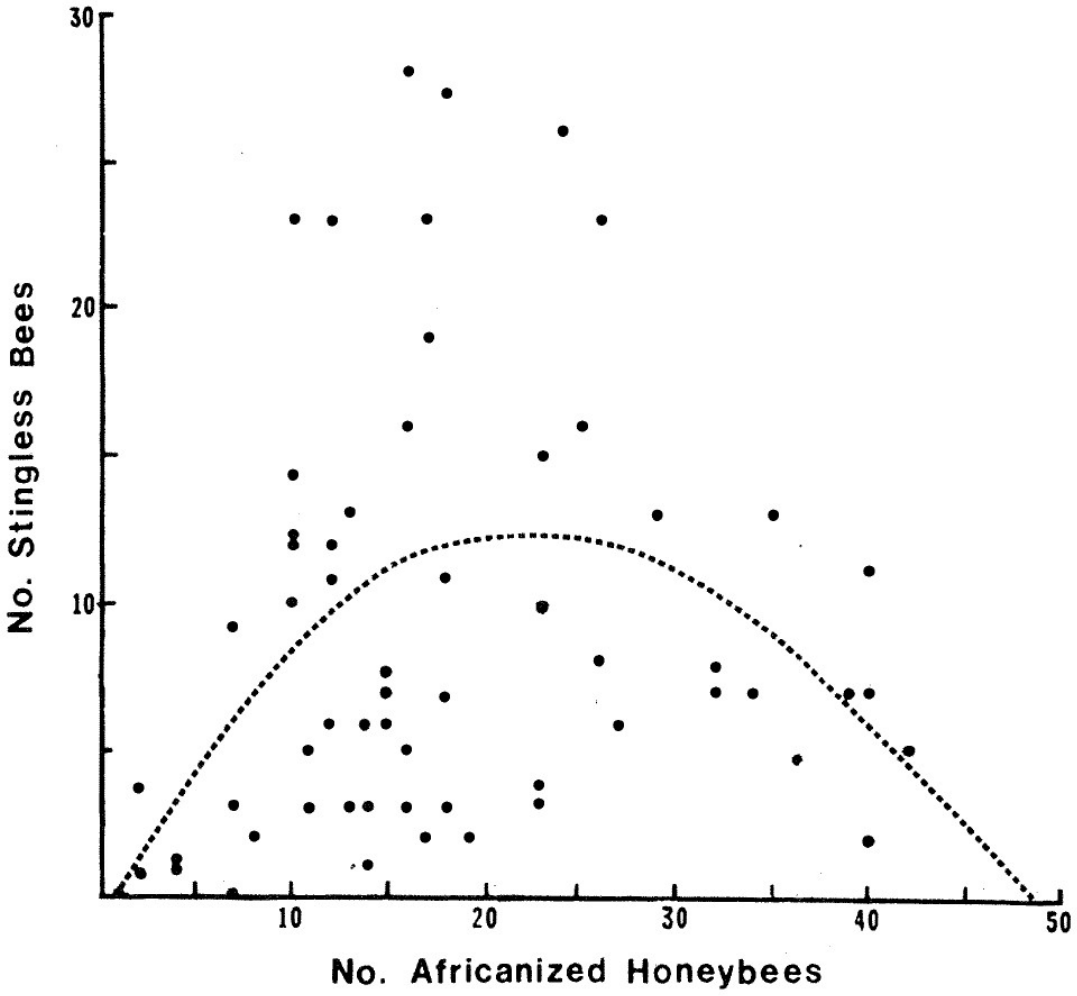


Fig. 1. The relations of Africanized and stingless (meliponine) bee abundances on flowering *Melochia villosa*. The dashed line is a quadratic polynomial (given by $y = -0.516 + 1.08x - 0.023x^2$) which gave the best fit to the points (*7*).

Caution: Bee careful

Just in case you missed it;

which
gave the best fit to the
points (7).

Check the text; he *really* means it

Stingless bees became less abundant
when Africanized honey bees increased
in numbers on *Melochia vellosa*

Caution: Bee careful

A stinging rebuke; (Hazen, *Science* Vol 201, 24th Nov 1978)

Curve-Fitting

The rather fanciful curve-fitting of Roubik (Reports, 15 Sept., p. 1030, Fig. 1) has prompted me to propose an alternative interpretation of his data (see below).

ROBERT M. HAZEN

Geophysical Laboratory,
Carnegie Institution of Washington,
Washington, D.C. 20018

Caution: Bee careful

- The fitted line was

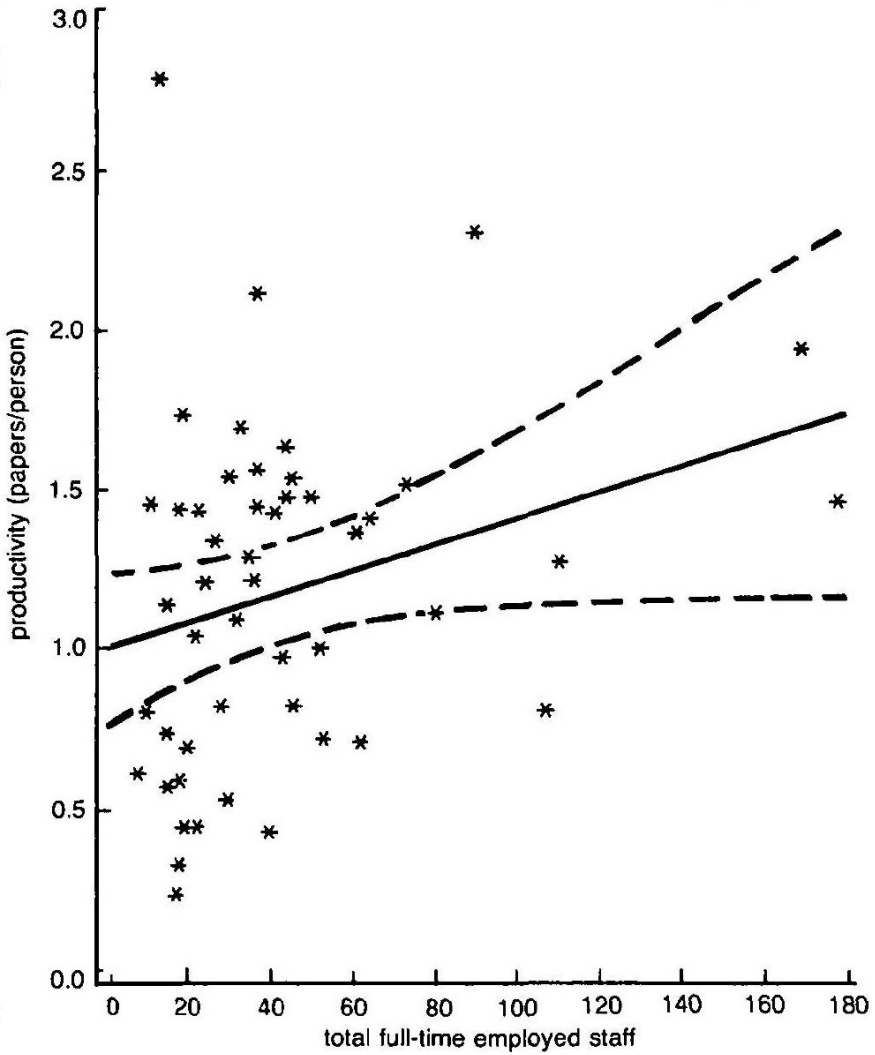
$$-0.516 + 1.08x - 0.023x^2.$$

- The estimates are *easy* to justify, but...
- The 'significance' of the fit, compared to a linear relationship, relies on an assumed Normal distribution
- 'Classical' methods *always* made this assumption (and with lab science data it may be appropriate)
- Modern methods allow us to make less assumptions – appropriate for lots of biostat work
- ... turns out 'messy' data need **far bigger** samples before anyone will believe your quadratic. Don't base conclusions on getting $p < 0.05$, once

So should you take stats problems to *physicists*? ...

Caution: physicists at work

'Is bigger really better?' (Physics World, 1989)



In British physics labs, Hicks and Skea found *some* evidence that papers/person increased with size of lab (left).

The fitted line gives $p = 0.049$ – can you guess the conclusions?

(Oxford and Cambridge are the two biggest labs)

Caution: physicists at work

Another *waspish* reply...

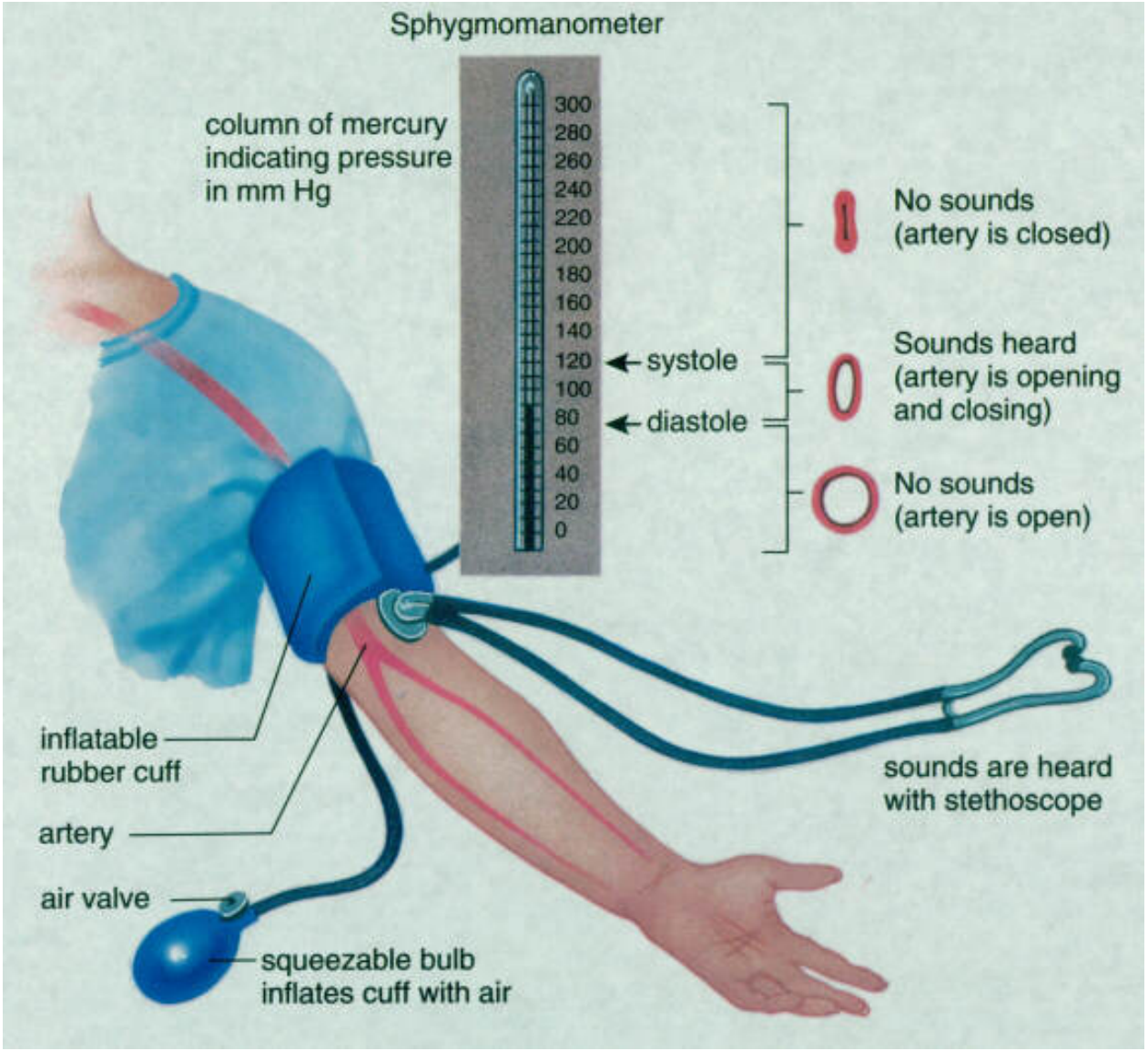
Fit for nothing?

The classic paper by D W Roubik (1978 'Competitive interactions between neotropical pollinators and Africanised honeybees' *Science* **201** 1030), with its skilful demonstration of computer processing of experimental data, has been superseded and improved. A new exemplar has been published with pride by *Physics World*. In the spirit of Robert M Hazen (1987 *Science* **202** 823), may I submit an alternative fit (see figure) to the data reported by Hicks and Skea (December p34).

Caution: physicists at work

- A linear trend is a **reasonable thing to record**, even if the 'truth' isn't a straight line
- For observational data you **certainly** want to note it
- Report the point estimate and 95% interval, not just '$p < 0.05$'
- Extreme covariates (Oxford, Cambridge) leads to high *leverage*; taking them out leads to a **very different answer**. Report this! Be honest! (Hicks and Skea did)
- The Hicks & Skea analysis is not *that* bad – avoid writing snotty letters unless you are an expert...

Caution: Hypertension can be risky



Caution: Hypertension can be risky

A serious example from a serious journal;

The Relation Between Pulse Pressure and Cardiovascular Mortality in 12 763 Middle-aged Men From Various Parts of the World

A 25-Year Follow-up of the Seven Countries Study

Demosthenes B. Panagiotakos, PhD; Daan Kromhout, PhD; Alessandro Menotti, MD, PhD; Christina Chrysohoou, MD, PhD; Anastasios Dontas, MD; Christos Pitsavos, MD, PhD; Hisashi Adachi, MD, PhD; Henry Blackburn, MD, PhD; Srecko Nedeljkovic, MD, PhD; Aulikki Nissinen, MD, PhD

ARCH INTERN MED/VOL 165, OCT 10, 2005

Pulse/mid/mean pressure; which **one** predicts best? A reasonable question for this data (although systolic and diastolic **together** are superior)

Caution: Hypertension can be risky

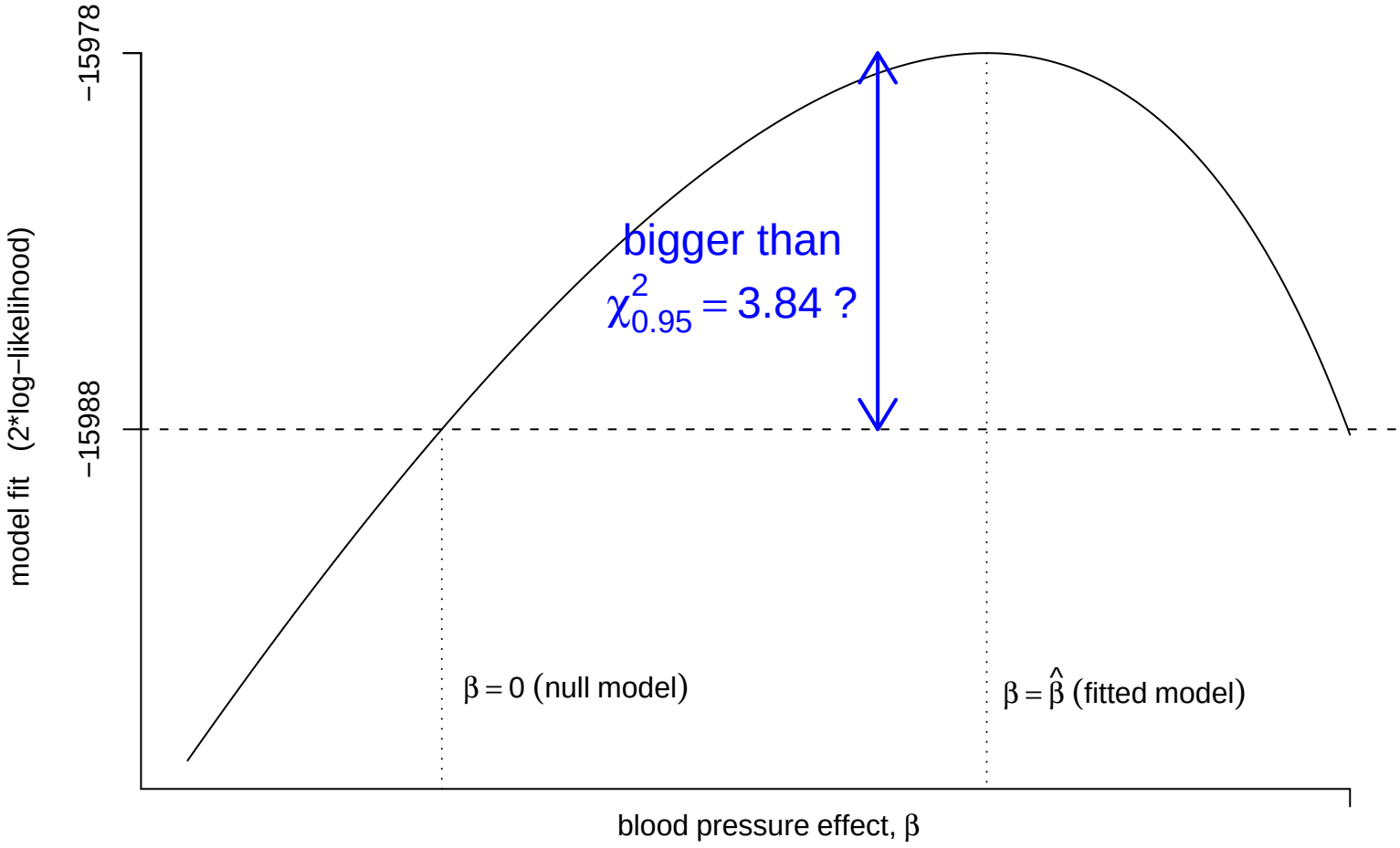
Here's what Panagiotakos et al reported;

Results: pulse pressure and diastolic and systolic blood pressures were the best predictors for CVD death, followed by mean and mid blood pressures.

Conclusions: Pulse pressure followed by diastolic and systolic blood pressures were the best predictors for CVD mortality among other blood pressures,

Caution: Hypertension can be risky

How did they get there? By comparing 'model fit'. Compare the fit of a model where blood pressure does **nothing** to the model **suggested** by the data;



Caution: Hypertension can be risky

Here are the Panagiotakos 'χ^2' values;

Measurement	χ^2	Predictor
Pulse	15978	Best
Diastolic	15923	
Systolic	15874	
Mid	15642	
Mean	15652	Worst

These are;

i) *much* bigger that 3.84
ii)

Caution: Hypertension can be risky

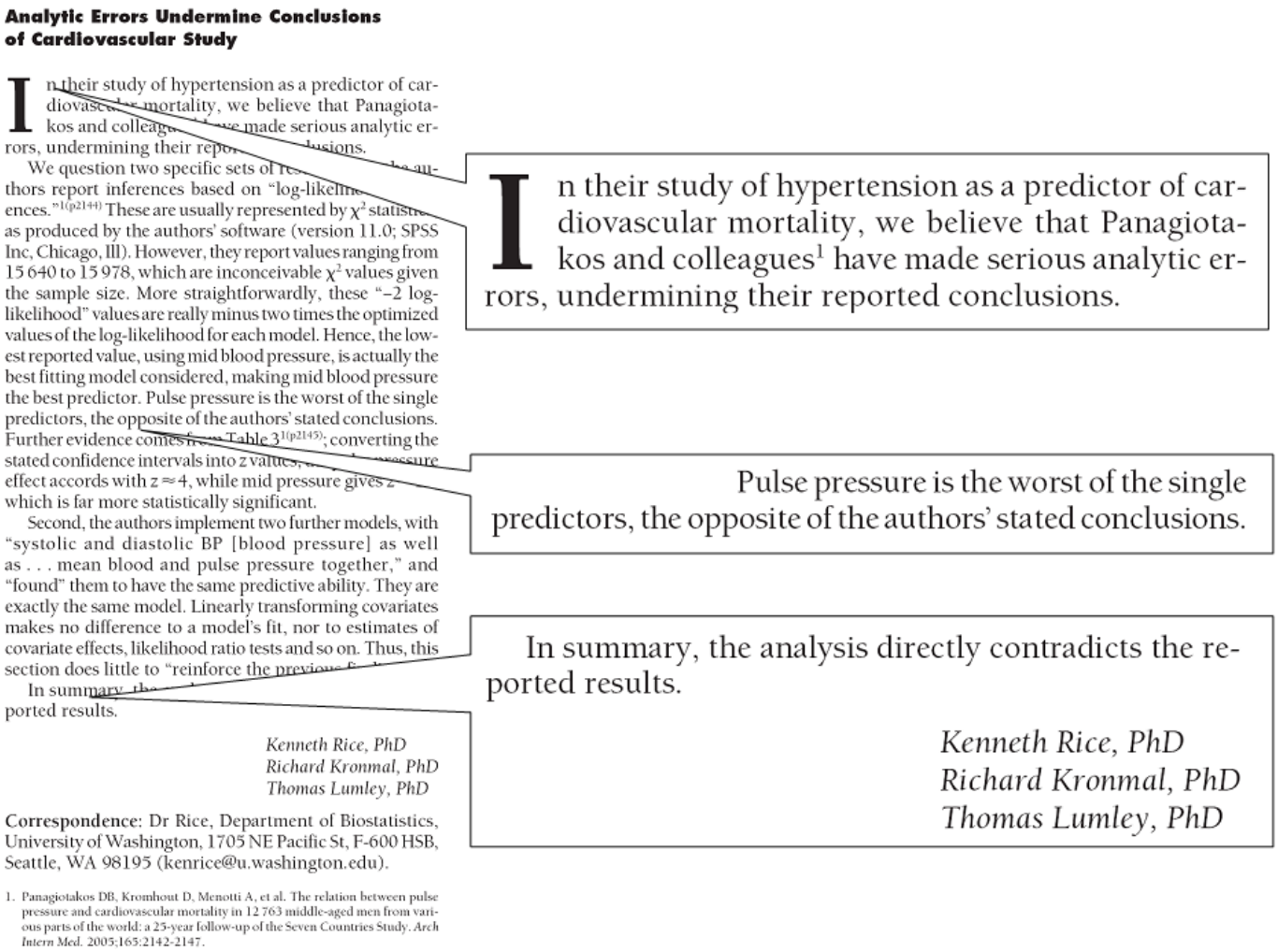
Could they have avoided this? Yes; look at Table 3;

Model	Hazard Ratio (95% Confidence Interval)
Pulse pressure (per 10 mm Hg)	1.22 (1.10-1.34)
Diastolic blood pressure (per 5 mm Hg)	1.16 (1.05-1.27)
Systolic blood pressure (per 10 mm Hg)	1.22 (1.10-1.35)
Mean blood pressure (per 10 mm Hg)	1.42 (1.38-1.47)
Mid blood pressure (per 10 mm Hg)	1.38 (1.30-1.41)

The pattern of significance is reversed.

Caution: Hypertension can be risky

A polite-but-snotty letter; (The paper was retracted)



Analytic Errors Undermine Conclusions of Cardiovascular Study

In their study of hypertension as a predictor of cardiovascular mortality, we believe that Panagiotakos and colleagues made serious analytic errors, undermining their reported conclusions.

We question two specific sets of results. The authors report inferences based on "log-likelihood differences."[1(p2144)] These are usually represented by χ^2 statistics as produced by the authors' software (version 11.0; SPSS Inc, Chicago, Ill). However, they report values ranging from 15 640 to 15 978, which are inconceivable χ^2 values given the sample size. More straightforwardly, these "−2 log-likelihood" values are really minus two times the optimized values of the log-likelihood for each model. Hence, the lowest reported value, using mid blood pressure, is actually the best fitting model considered, making mid blood pressure the best predictor. Pulse pressure is the worst of the single predictors, the opposite of the authors' stated conclusions. Further evidence comes from Table 3[1(p2145)]; converting the stated confidence intervals into z values, ... pressure effect accords with $z \approx 4$, while mid pressure gives z ... which is far more statistically significant.

Second, the authors implement two further models, with "systolic and diastolic BP [blood pressure] as well as . . . mean blood and pulse pressure together," and "found" them to have the same predictive ability. They are exactly the same model. Linearly transforming covariates makes no difference to a model's fit, nor to estimates of covariate effects, likelihood ratio tests and so on. Thus, this section does little to "reinforce the previous ...

In summary, the ... ported results.

Kenneth Rice, PhD
Richard Kronmal, PhD
Thomas Lumley, PhD

Correspondence: Dr Rice, Department of Biostatistics, University of Washington, 1705 NE Pacific St, F-600 HSB, Seattle, WA 98195 (kenrice@u.washington.edu).

1. Panagiotakos DB, Kromhout D, Menotti A, et al. The relation between pulse pressure and cardiovascular mortality in 12 763 middle-aged men from various parts of the world: a 25-year follow-up of the Seven Countries Study. *Arch Intern Med*. 2005;165:2142-2147.

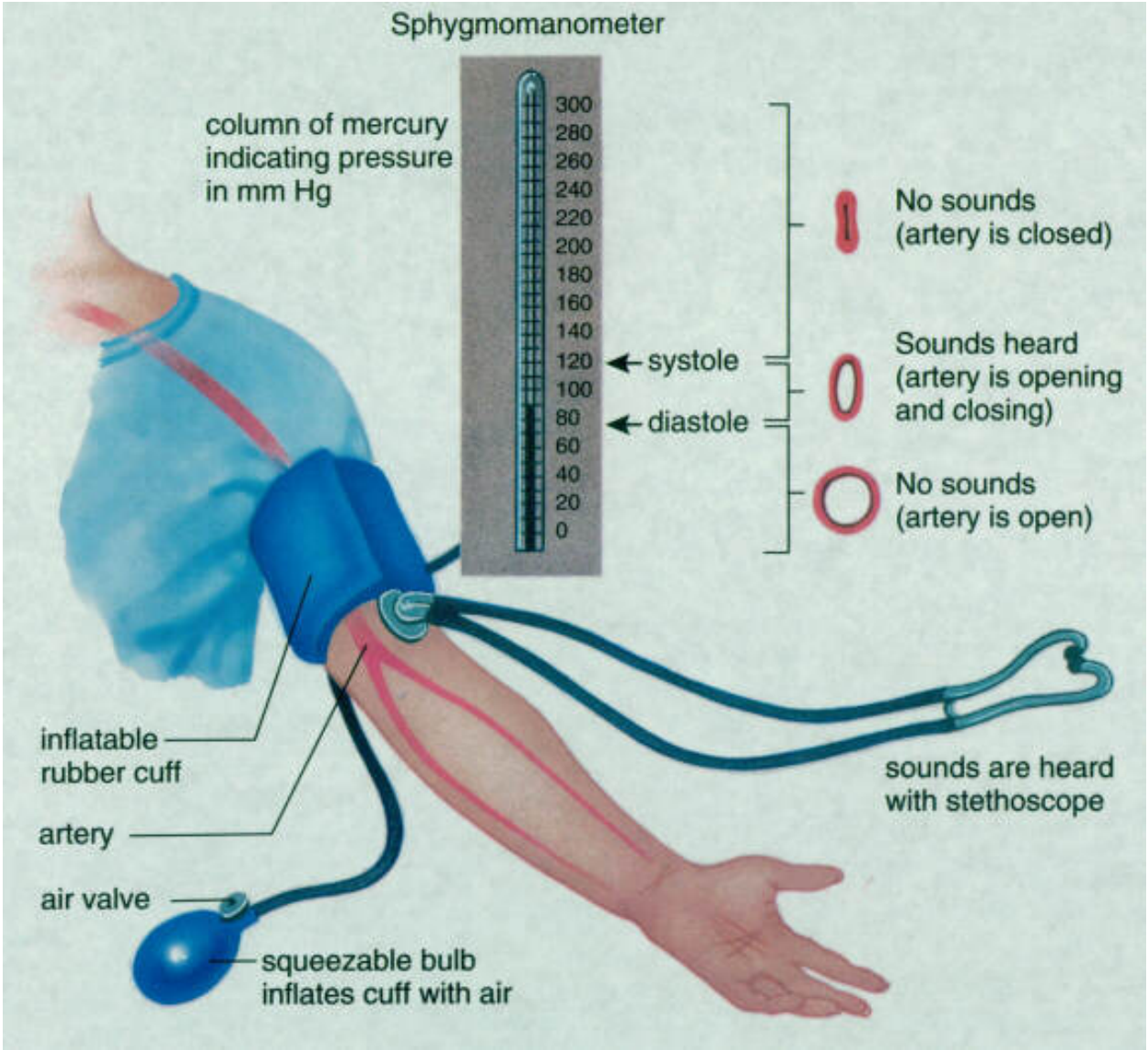
In their study of hypertension as a predictor of cardiovascular mortality, we believe that Panagiotakos and colleagues[1] have made serious analytic errors, undermining their reported conclusions.

Pulse pressure is the worst of the single predictors, the opposite of the authors' stated conclusions.

In summary, the analysis directly contradicts the reported results.

Kenneth Rice, PhD
Richard Kronmal, PhD
Thomas Lumley, PhD

Caution: The TROPHY goes to...

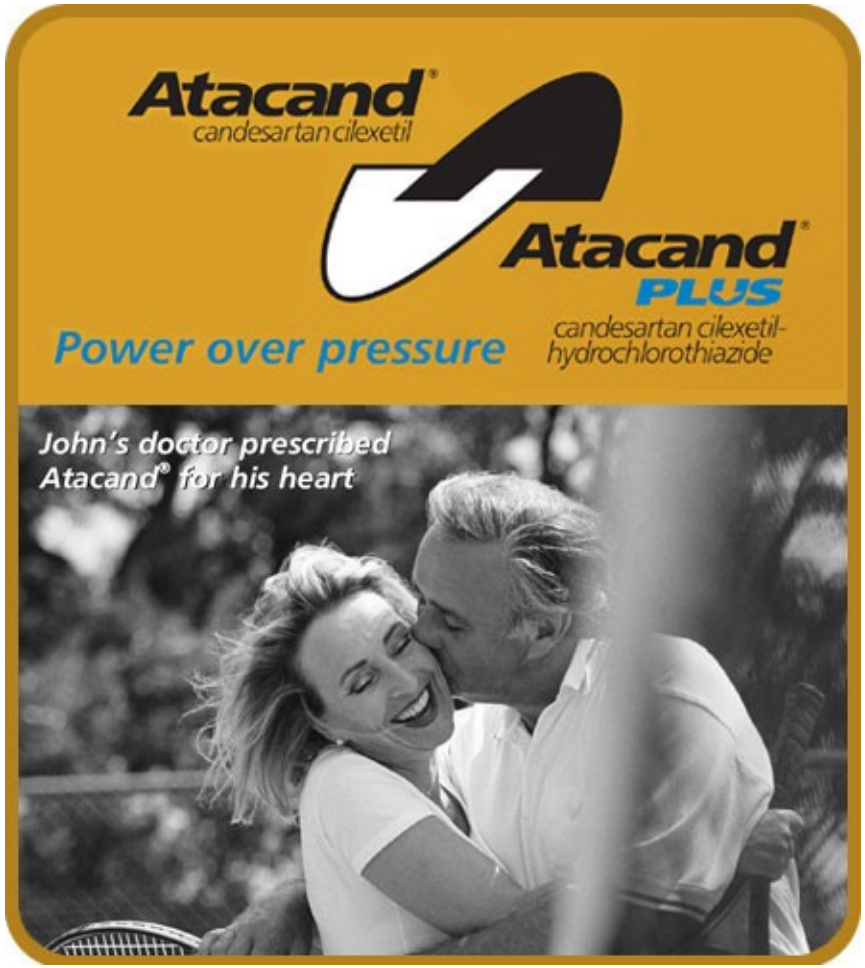


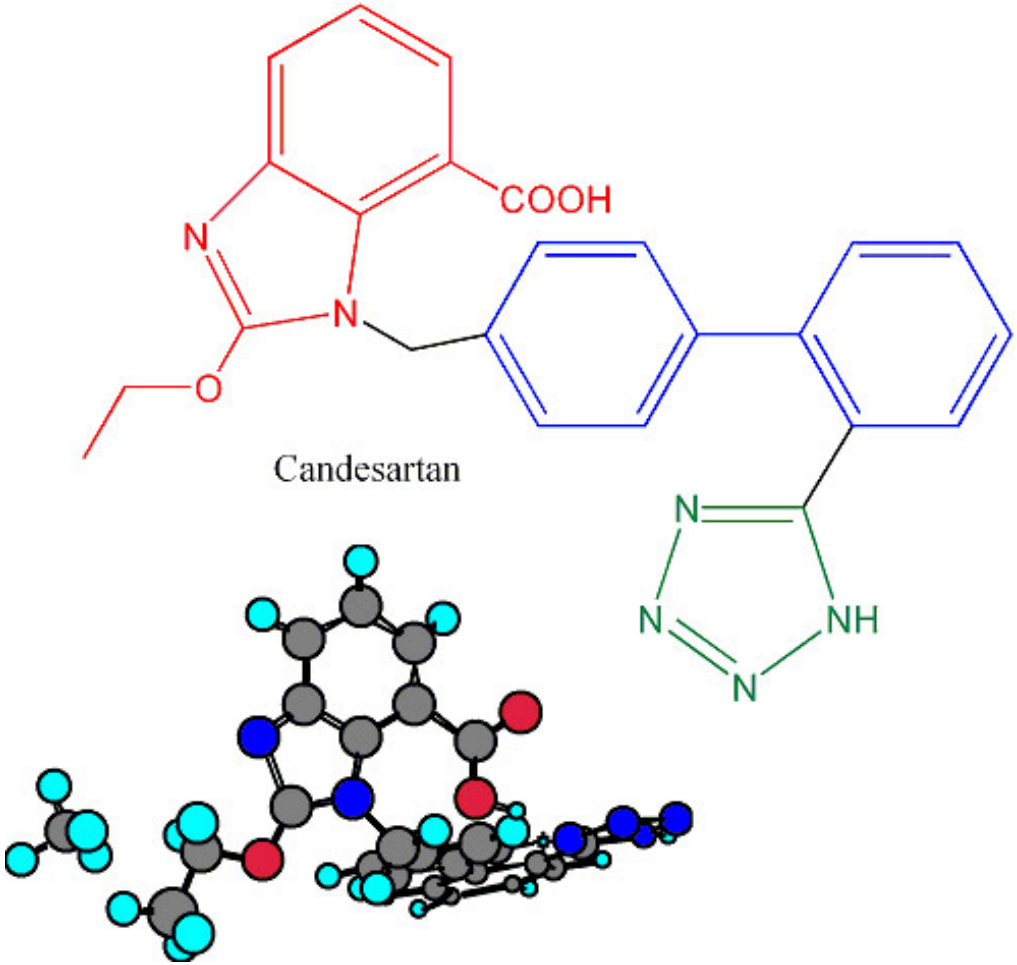
Blood pressure

Why do we tell people to eat less salt?

- About 30% of adults have **High Blood Pressure** – African Americans have even more
- Risk factor for stroke, heart attack, heart failure, kidney disease... death
- Definitions of HBP involve thresholds, e.g. SBP>140mmHg
- 'Environment' matters (e.g. salt intake) but known to have a genetic component...*partly* explained by Rice *et al* 2009

Antihypertensive drugs





Candesartan

Antihypertensive drugs

- These work! Lowering BP by about 10 mmHg (systolic), 5 mmHg (diastolic)

- Sales have grown to ∼$35 billion per year ... as the 'threshold' for HBP gets lower!

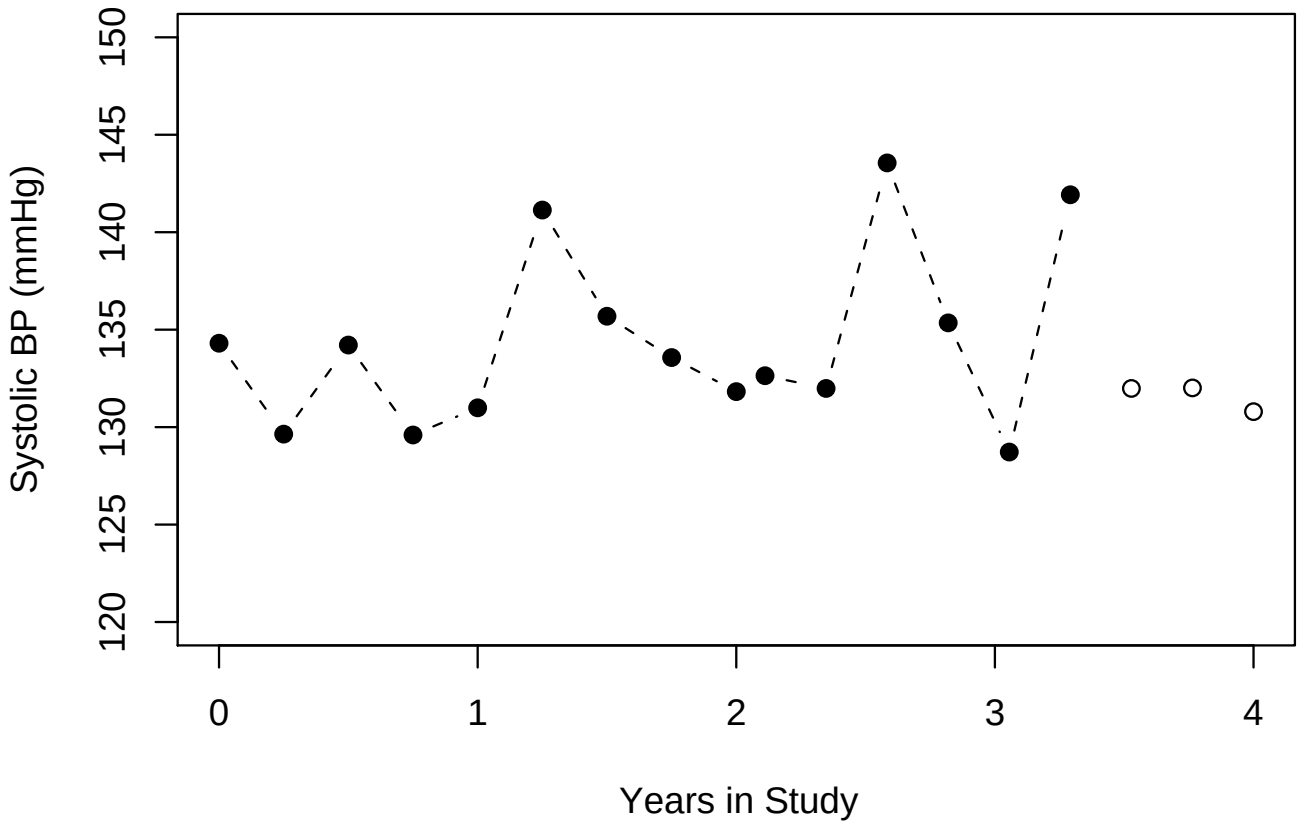
- Once prescribed, typically stay on BP drugs **for life**

- Long-term behavior is of interest – is there a 'carryover' effect of drugs?

AstraZeneca's **TR**ial **O**f **P**reventing **HY**pertension (TROPHY) examined the carryover effect of candesartan, an accepted BP drug

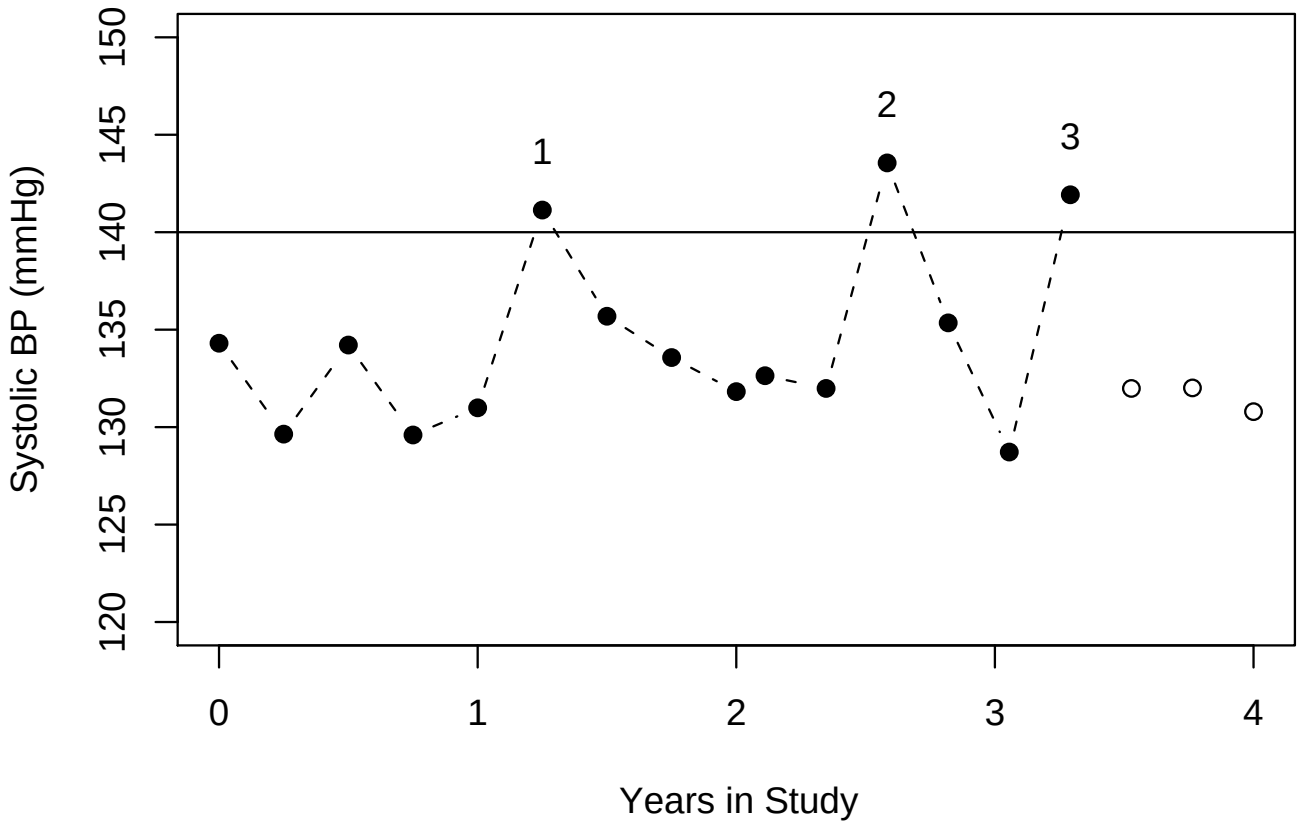
TROPHY's endpoint

Measure blood pressure 18 times over two years ($>$ 140 is bad)



TROPHY's endpoint

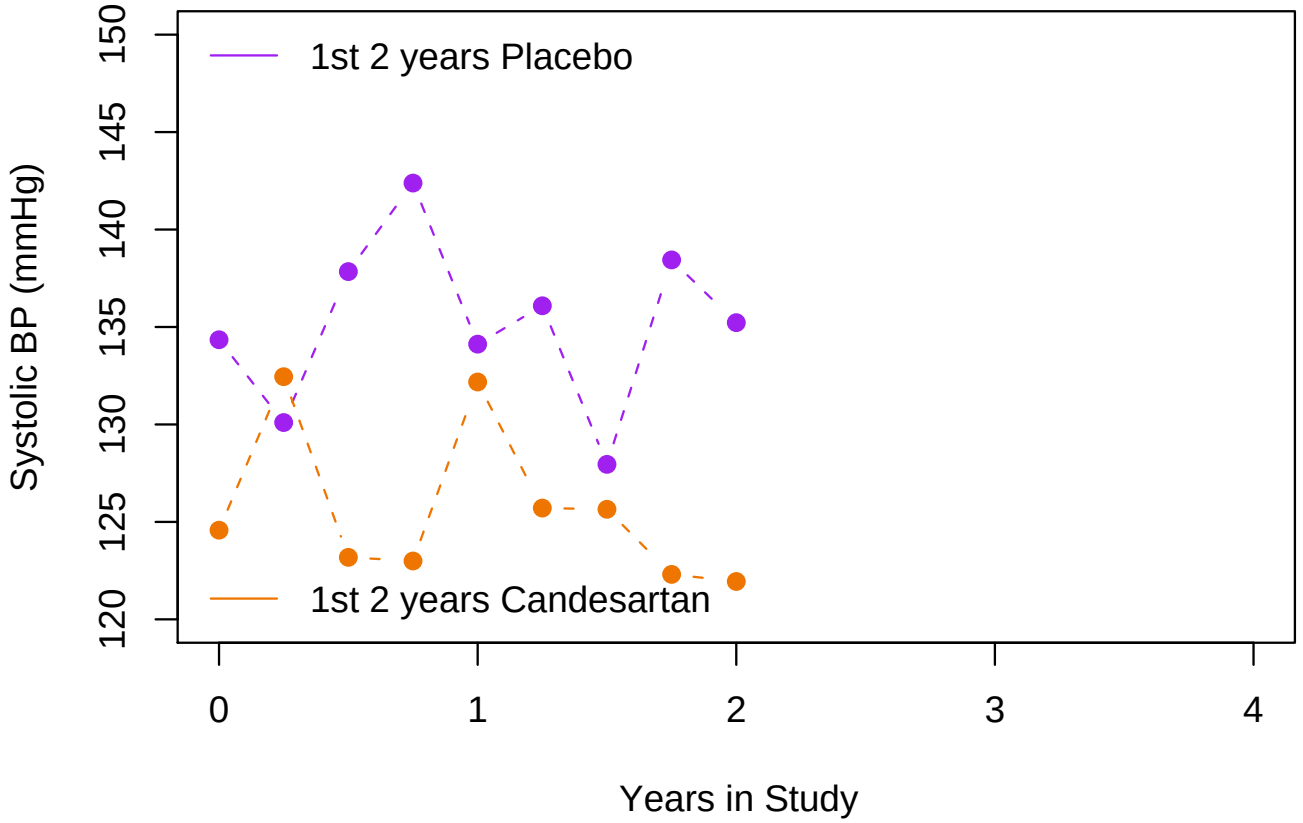
Measure blood pressure 18 times over two years (> 140 is bad)



'Three strikes and you're out' – i.e. an **incident case**, and put on BP drugs for life

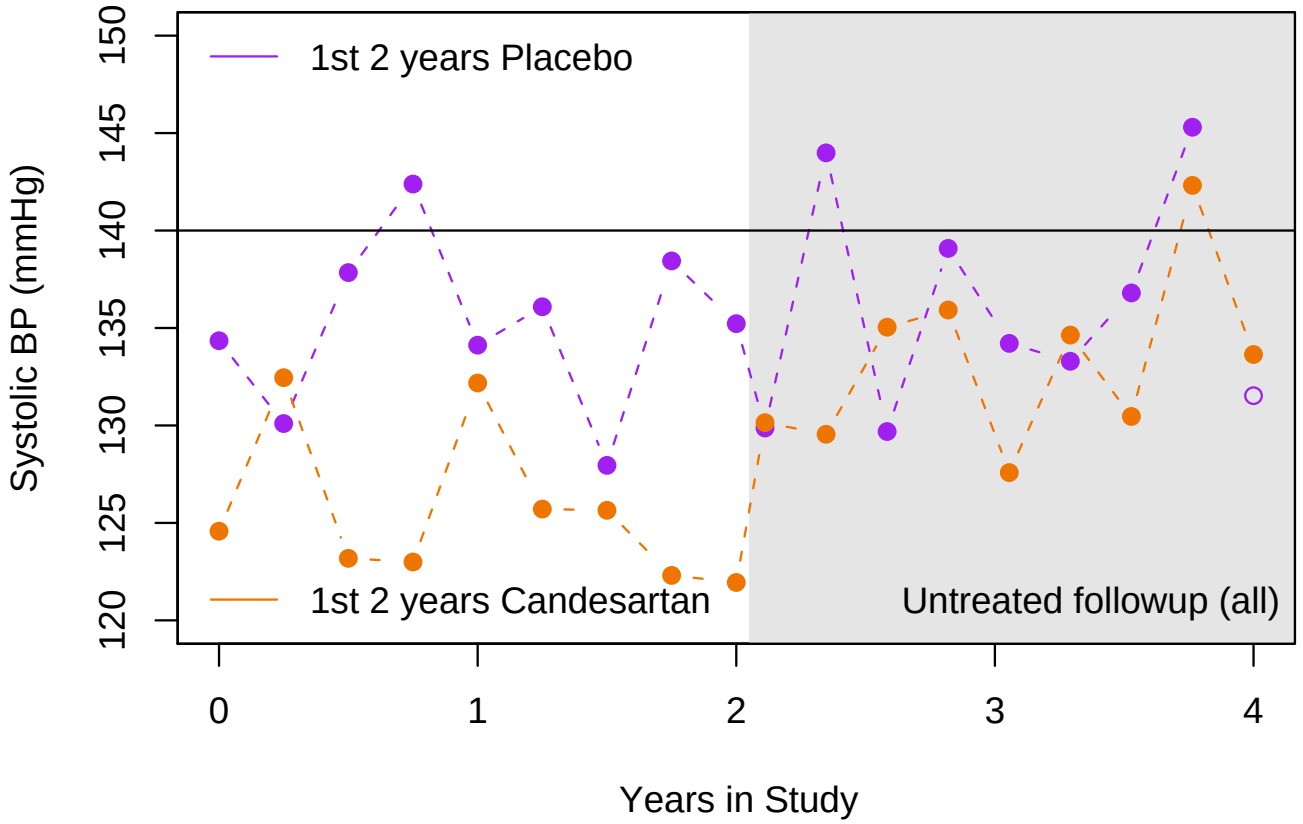
TROPHY's carryover effect

For the first two years, randomize patients to candesartan or placebo;



TROPHY's carryover effect

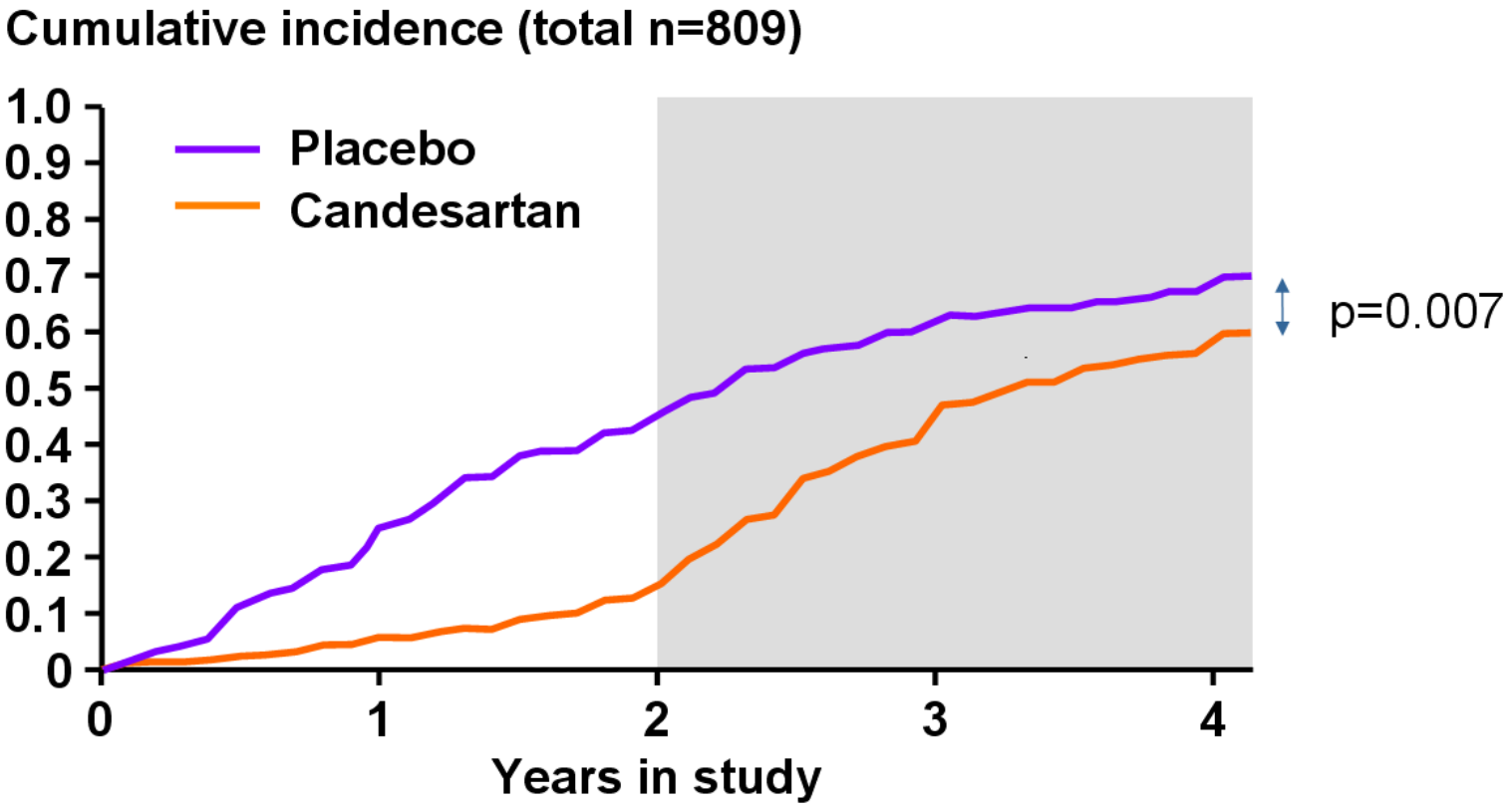
For the first two years, randomize patients to candesartan or placebo;



... then two years of non-treatment. 'Carryover' means lower line would **stay low**.

TROPHY – what happened?

On placebo/candesartan, what fraction got three HBP ‘strikes’?



Source: `www.atacand.com` ©AstraZeneca

... so there’s a carryover effect, right?

TROPHY – what happened?

The NEW ENGLAND JOURNAL of MEDICINE

"...the effect of active treatment on delaying the onset of hypertension can extend up to 2 years after the discontinuation of treatment."
Julius et al, NEJM (2006) 354:1685–1697

A high-impact paper, with two *fantastic* messages;

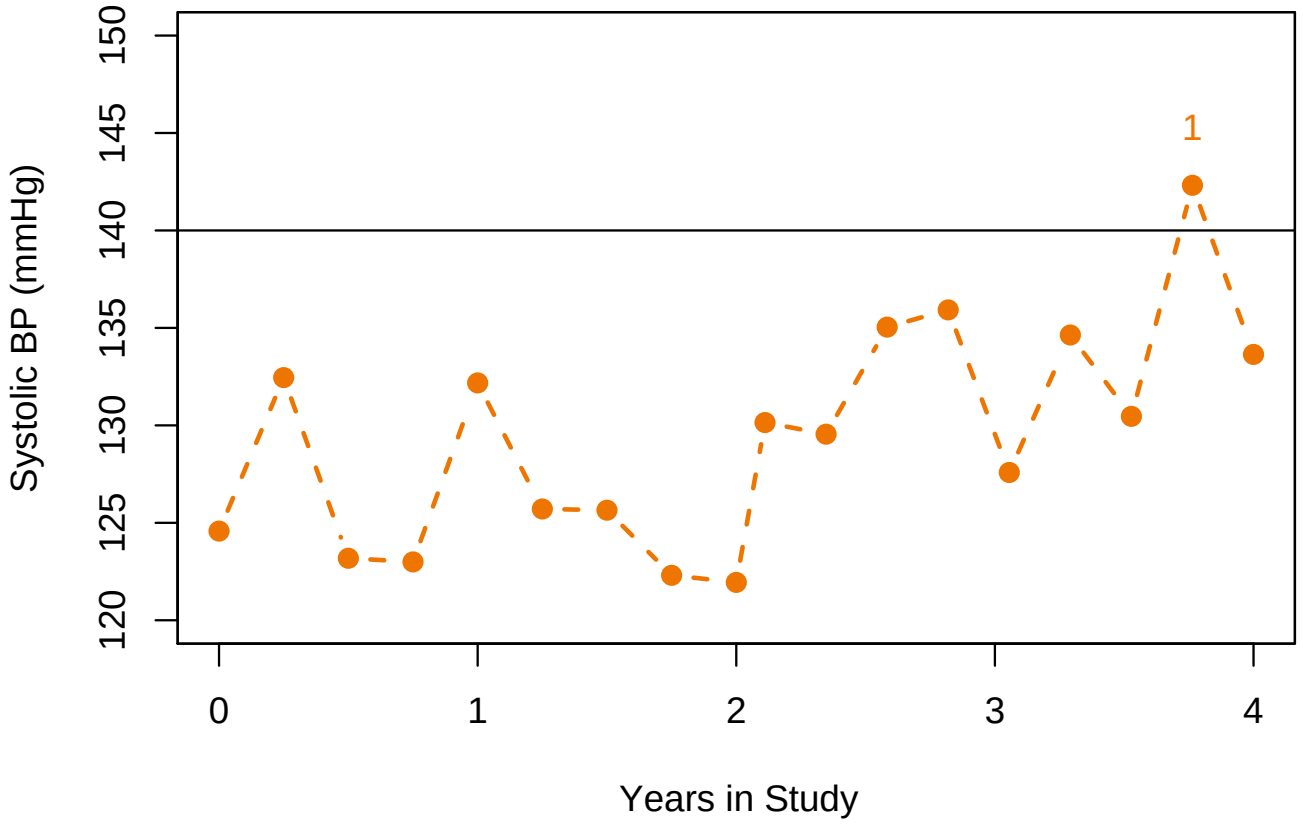
- Candesartan carries over (*"modestly"*) for ∼2 years
  ...a *huge* effect for a drug **you are not taking!**
- We should treat 'prehypertensives' now
  ...instead of **treating them later?**

TROPHY – what do you want to know?

- TROPHY **does** gives a fair comparison for the 2+2 years, i.e. **treatment + carryover**
- The null hypothesis is of **no treatment effect, no carryover effect** – and we rejected it
- But first two 'treatment' years are **not interesting** – we *already know* Candesartan works then
- 'Carryover' is what got TROPHY into NEJM
- Is TROPHY a fair comparison in years 3&4?

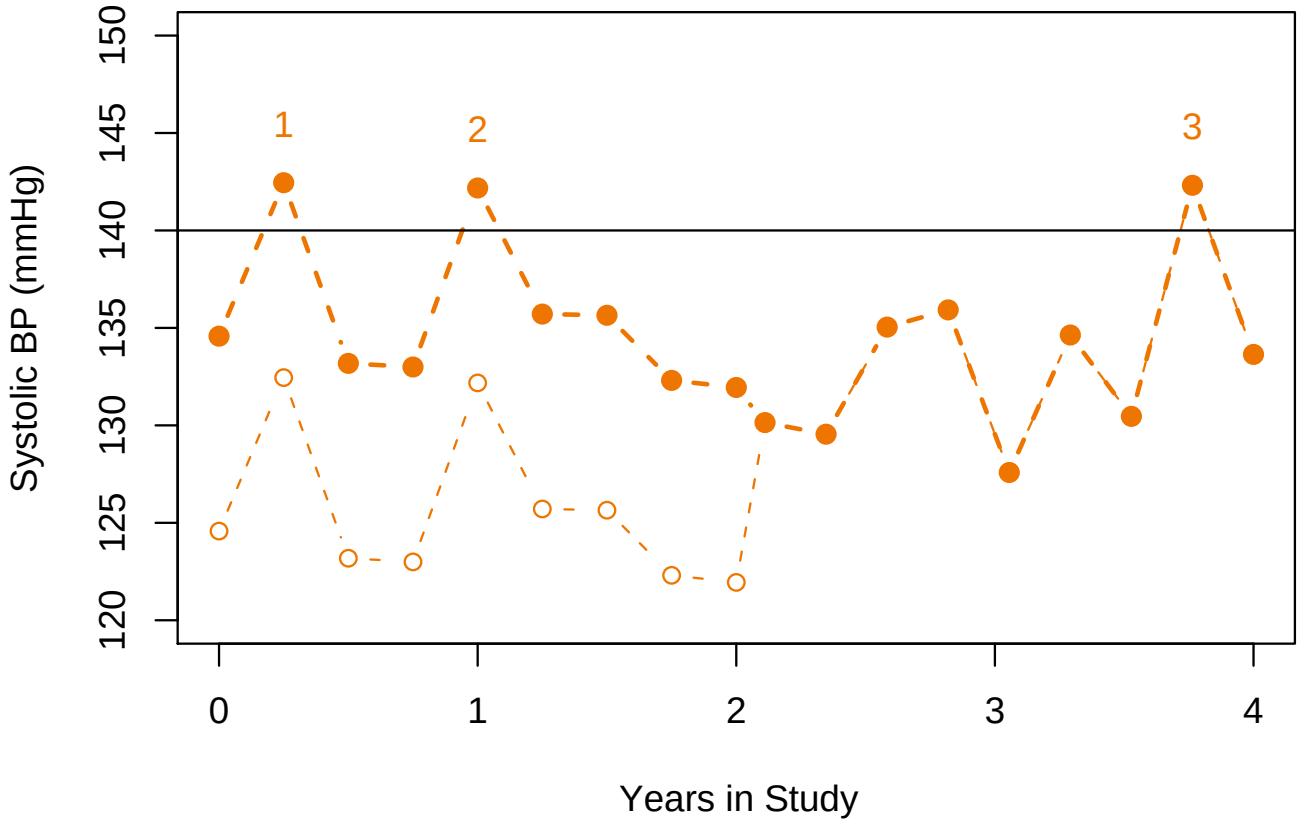
TROPHY – a thought experiment

Take someone on Candesartan; what happens to them on placebo, with **no carryover?**



TROPHY – a thought experiment

Take someone on Candesartan; what happens to them on placebo, with **no carryover?**



They 'become' incident – because of years 1&2, **not** carryover

TROPHY – a simulation

- Being on placebo in years 1&2 gives a 'head start' on the 3 HBP 'strikes' – **even with zero carryover**

- TROPHY's analysis ignored this – but does it matter? Is the problem actually serious?

- This is a hard question – but is easily answered with simulations

Let's 'run the universe again,' without carryover, and try TROPHY's analysis...

TROPHY – a simulation

Let's 'run the universe again,' without carryover...

How to simulate? From prior studies, we know a **lot** about blood pressure behavior;

- Starting level ($\sim$130 mmHg, in TROPHY)
- Increase with age (1 mmHG per year)
- Measurement error ($\pm$6 mmHg at each measurement)
- Effect of treatment (drop of 10 mmHg, when on treatment)

With 10 mmHG treatment effect, **zero** carryover effect, what p-values should we expect under TROPHY's test of **treatment + carryover**?

Some TROPHYs to take home

- Same phenomenon occurs in diabetes studies, and other 'surrogate endpoints'

- Post-publication review is **very useful and very important** – but may not make you popular!

- Measurement error issues (and much else) are **non-intuitive** – statisticians help by figuring out **what the data mean**

Caution: Mind the (gender) gap

Some high-profile work in human physiology; (Tatem et al)

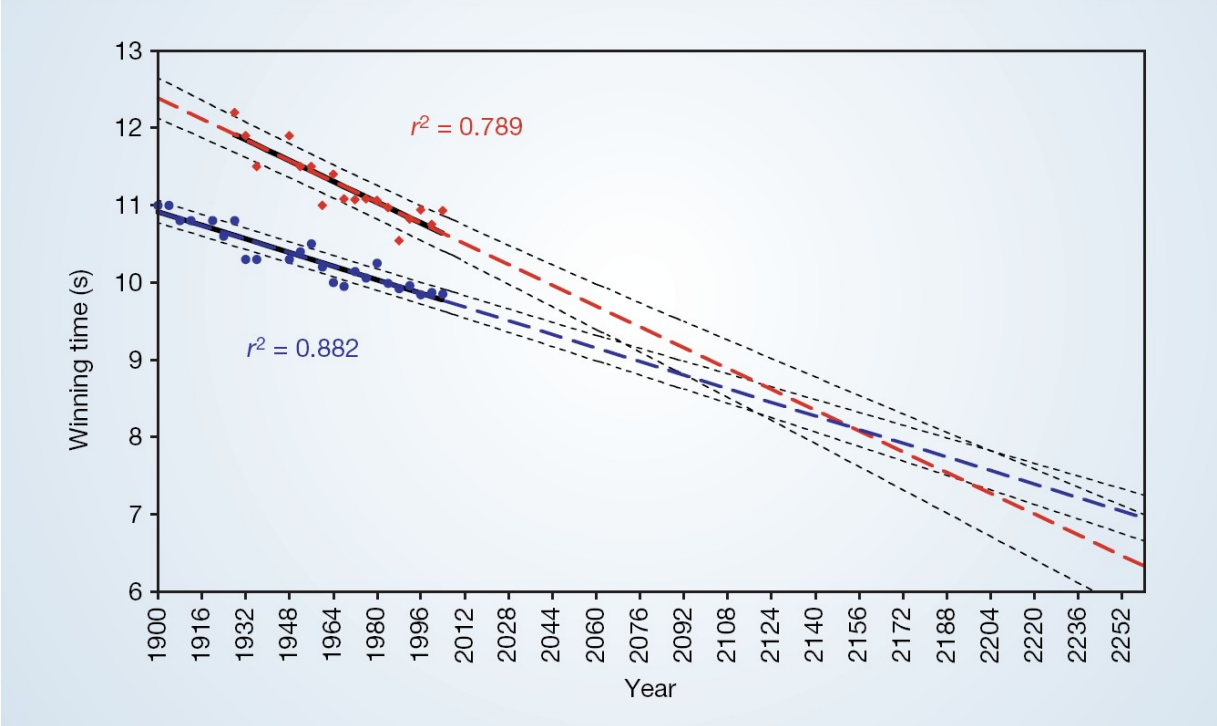
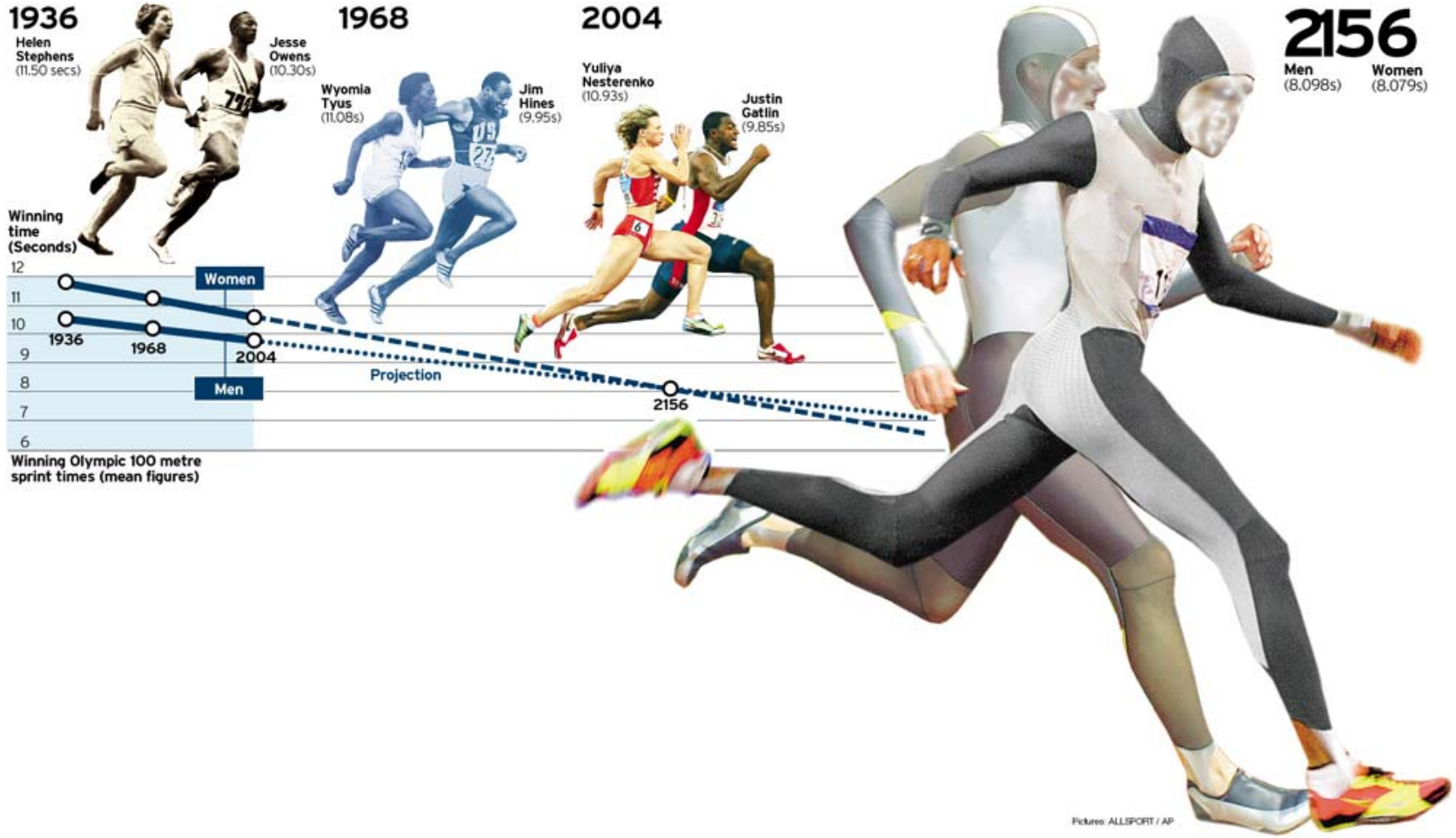


Figure 1 The winning Olympic 100-metre sprint times for men (blue points) and women (red points), with superimposed best-fit linear regression lines (solid black lines) and coefficients of determination. The regression lines are extrapolated (broken blue and red lines for men and women, respectively) and 95% confidence intervals (dotted black lines) based on the available points are superimposed. The projections intersect just before the 2156 Olympics, when the winning women's 100-metre sprint time of 8.079 s will be faster than the men's at 8.098 s.

NATURE | VOL431 | 30 SEPTEMBER 2004 | www.nature.com/nature

Caution: Mind the (gender) gap

The press *loved* it; ‘Those wacky scientists have shown...’



Caution: Mind the (gender) gap

Another politely-worded letter...

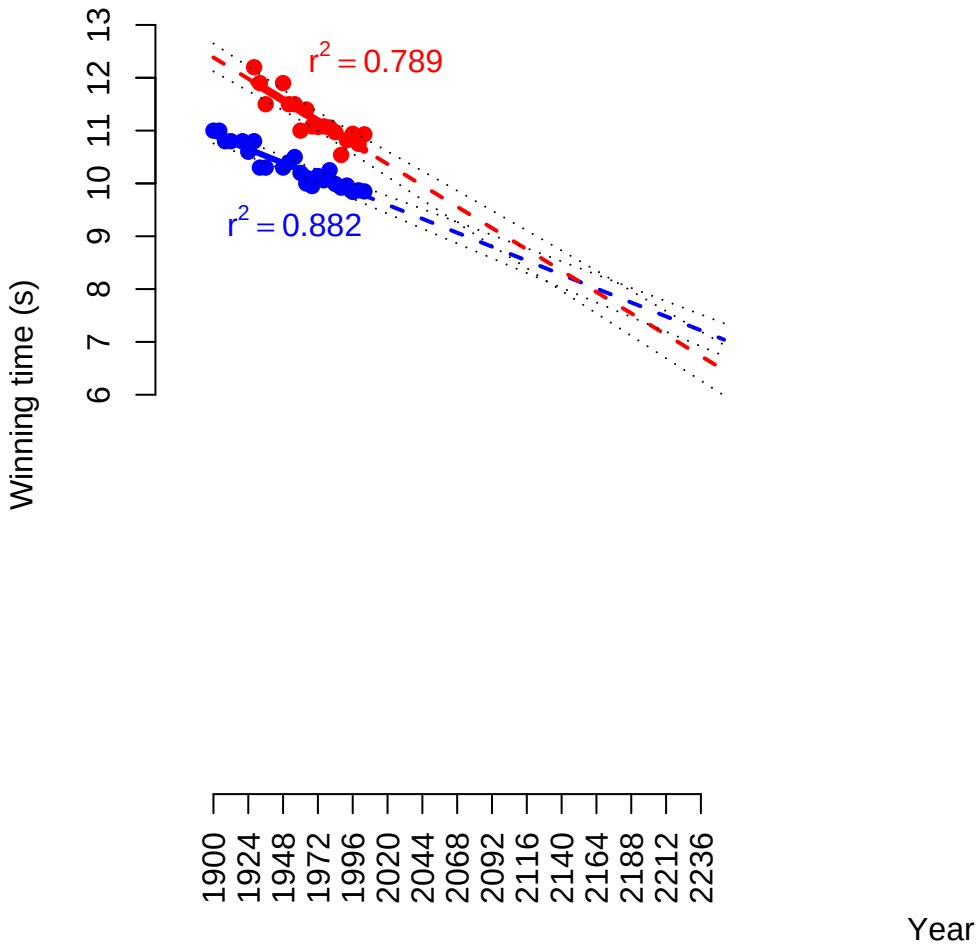
Sprint research runs into a credibility gap

Sir — A. J. Tatem and colleagues calculate that women may outsprint men by the middle of the twenty-second century (*Nature* **431,** 525; 2004). They omit to mention, however, that (according to their analysis) a far more interesting race should occur in about 2636,

NATURE | VOL 432 | 11 NOVEMBER 2004 | www.nature.com/nature

Caution: Mind the (gender) gap

Let's look at that extrapolation again;



Caution: Mind the (gender) gap

- The original was a *light-hearted* piece, albeit about a serious hypothesis
- A group of high school biology students wrote a (great) statistical review; Nature published that too
- 'Extrapolation' is estimating quantities far from those measured in your dataset. It is common and useful, but requires **great care**, and **sensitivity analysis**

Take precautions with statistics, as you would do elsewhere...

Summary

- (Bio)statisticians are most help with **subtle** problems
- We want smart people (i.e. **you**) who will think *hard*
- UW courses reflect this; as much scientific **thought** as statistical **computation**– a rare combination.
- Math is important, but it's not the whole story

For a flavor of consulting, and more in-depth work;

`www.biostat.washington.edu`
`faculty.washington.edu/~kenrice`