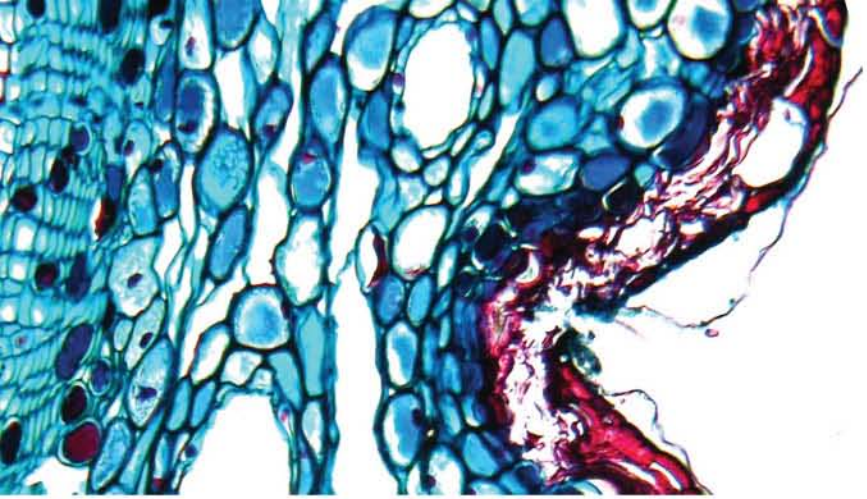
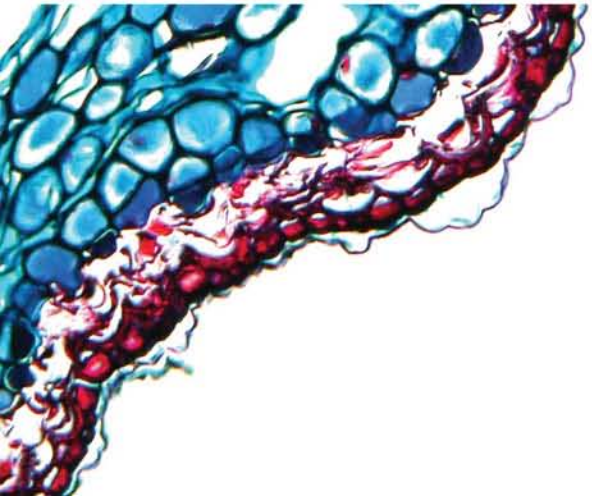


QUINTILES®

Navigating the new health

Imperial College LTTC Course
10MAR2010
Andy Garrett

A personal view on how statistics is used in practice in the pharmaceutical industry



clinical | commercial | consulting | capital

Content

- History of statistics in drug development and some basic concepts
- How to get into the profession
- Some simple illustrations of problems faced
- Two examples
  - Logistics regression, covariate adjustment and non-inferiority
  - Simpson’s paradox
  - RSS First in Man WP
- Current areas of statistics research in the industry

History of drug regulation

- Drug regulation driven primarily from US, but now International (ICH, 1991)
- 1906 Pure Food and Drug Act response to concerns adulteration and misbranding of food and drugs - focus was drug labelling not drug approval
- 1938 Food, Drug and Cosmetic Act required regulatory approval prior to marketing – proof of safety
- 1962 Kefauver-Harris Amendments added proof of efficacy
  - Requirement to submit *substantial evidence* to support regulatory approval and for this to originate from *adequate and well-controlled investigations*.
  - defining moment for statistics in drug development, made sound statistical methodology an integral part of the regulatory process
- Regulation as responses to notable tragic events - such as thalidomide related birth defects (1962 K-H Amendments) - or to growing concerns regarding activities – e.g. reluctance to conduct paediatric studies (2002 Best Pharmaceuticals for Children Act).

How do get into the profession

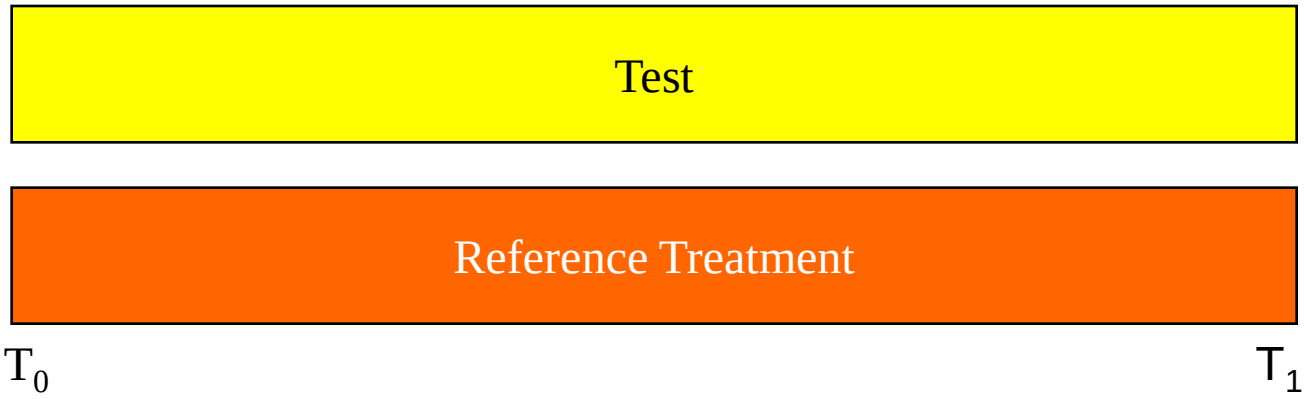
- Global
  - Travel, TCs! India, China

- Teamwork
  - Communication
  - Organised

- Masters / PhD
  - Continuing professional development

- Pharmaceutical / biotech
- CROs (Pharmaceutical Services) – e.g. Quintiles
- Academic/Research – e.g. MRC

Basic randomised, controlled trial



- RA Fisher developed randomised experimentation with Bradford Hill applying to clinical research in the 1950's

-Over all randomisations, treatment groups will be balanced with respect to both known & unknown factors that influence outcome.
- If exclude patients may introduce bias (Intent to treat principle)

- Estimate the treatment difference at T_1 , construct CI to give range of plausible values. Most likely this will involve modelling (covariate adjustment)

Some simple illustrations – because they have an impact!

- Change from BL
- % change from BL
- Dichotomisation of endpoints

Senn & Julious (2009)

Use of baseline to construct change scores

Change scored is $D = Y - X$
where Y is outcome variable and X is the same variable at baseline

T-test to compare two treatments (T) is essentially
$(Y - X) = \alpha + \beta_2 T$

Re-arrange to get
$Y = \alpha + X + \beta_2 T$

ANCOVA is generally more efficient (i.e. more powerful)

$Y = \alpha + \beta_1 X + \beta_2 T$ (In t-test, forcing $\beta_1 = 1$)

Often one sees Change from baseline, with baseline as a covariate

$(Y - X) = \alpha + \beta_3 X + \beta_2 T$

$Y = \alpha + (\beta_3 + 1) X + \beta_2 T)$ (Identical parametrisation of treatment, but $\beta_1 = (\beta_3 + 1)$)

Key point is that baseline should be fitted as a covariate since the relative efficiency of (Y-X) versus ANCOVA is $(1+ \rho)/2$, where ρ is the correlation between Y and X

Percentage change from baseline

100 x (Y – X) / X

Same as 100 x (Y/X – 1), so working part is simply Y/X

Ratio unlikely to be normal, even if Y and X are normal

- Tends to be more of an issue when X is small and changes are large
- Data will be approx Normal if means are large compared to SD

Ratio are not good candidates for parametric analysis therefore

Typically take Logs instead

$\log(Y) = \alpha + \beta_1 \log(X) + \beta_2 T$

Note cannot take logs of negative numbers or zero

Key point: do not analyse data as percentage change from baseline

The loss in creating artificial dichotomies

Dichotomisation of normal data

Pitman efficiency of sign-test versus t-test is $(2/\pi) \approx 64\%$

- Assumes median split
- E.g. if abnormal is 2 SD from mean then relative efficiency is only 13%

Key finding: analyse on continuous scale but use dichotomisation to aid interpretation

If dichotomy is used then it is not possible to recover the full information provided by baseline (double whammy!)

A more detailed example – logistic regression, covariate adjustment & non-inferiority

Stratified binary data example (Balanced 2x2x2 table)

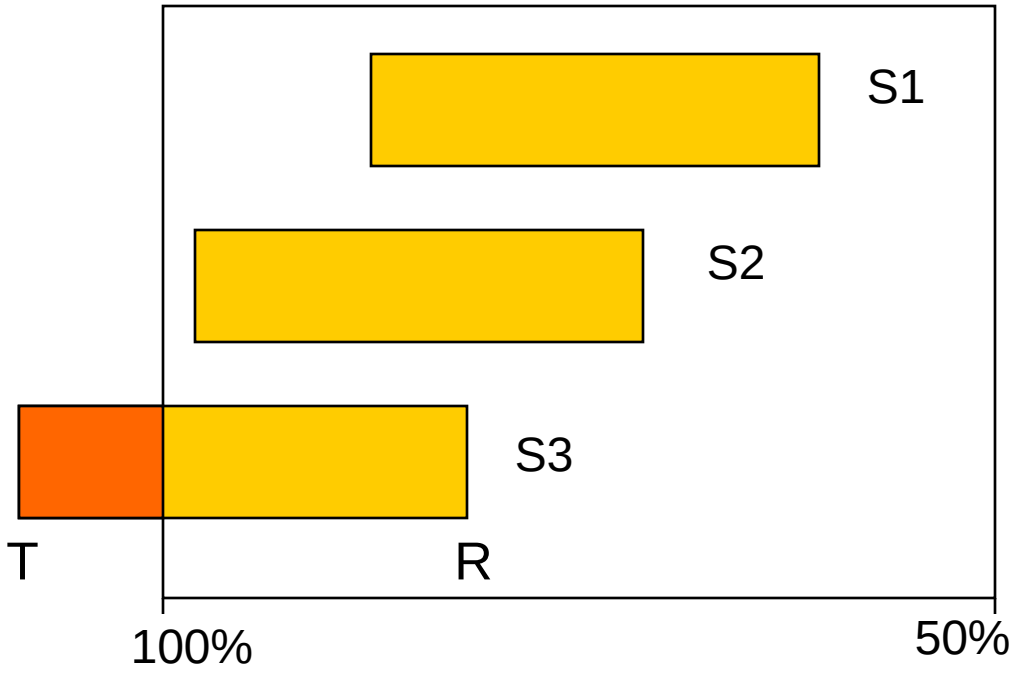
	Males Diff=15% OR=3	Females Diff=25% OR=3	Total Diff=20% OR=2.83
Test	**90/100 = 90%** odds 90/10 = 9	**75/100=75%** odds 75/25 = 3	**165/200=82.5%** odds 165/35 = 4.71
Reference	**75/100 =75%** odds 75/25 = 3	**50/100=50%** odds 50/50 = 1	**125/200=62.5%** odds 125/75 = 1.67

Basic principles: stratified analyses

- Choose scale of measurement
- Stratified model
  - Primary model, if stratified design (analyse as you design)
  - Estimate adjusted treatment difference
    - Estimate each *within stratum* treatment difference
    - Combine these estimates using a system of weights
  - Confidence interval
  - Significance test (p-value)

- Investigation of consistency of effect
  - treatment by factor interaction

Difference in Percentages



In stratum S3, the Reference response is too high such that the treatment difference cannot be consistent with those observed in strata S1 & S2

See Smith *et al* (1998) for stratified analyses for the difference in proportions. Also Koch & Carr (1990)

Logistic model

- Models the $\log_e$ (odds)
- β is a $\log_e$ (odds ratio)
- It follows that e^{β} is an odds ratio

$$\ln(odds) = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1.strata + \beta_2.treatment$$

Main effects model

$$\begin{bmatrix} \ln(odds_{1r}) \\ \ln(odds_{1t}) \\ \ln(odds_{2r}) \\ \ln(odds_{2t}) \end{bmatrix} = \begin{bmatrix} \alpha & & \\ \alpha & & +\beta_2 \\ \alpha & +\beta_1 & \\ \alpha & +\beta_1 & +\beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Log odds for reference cell

Increment for log odds for test treatment - that is, ln(OR)

Design matrix

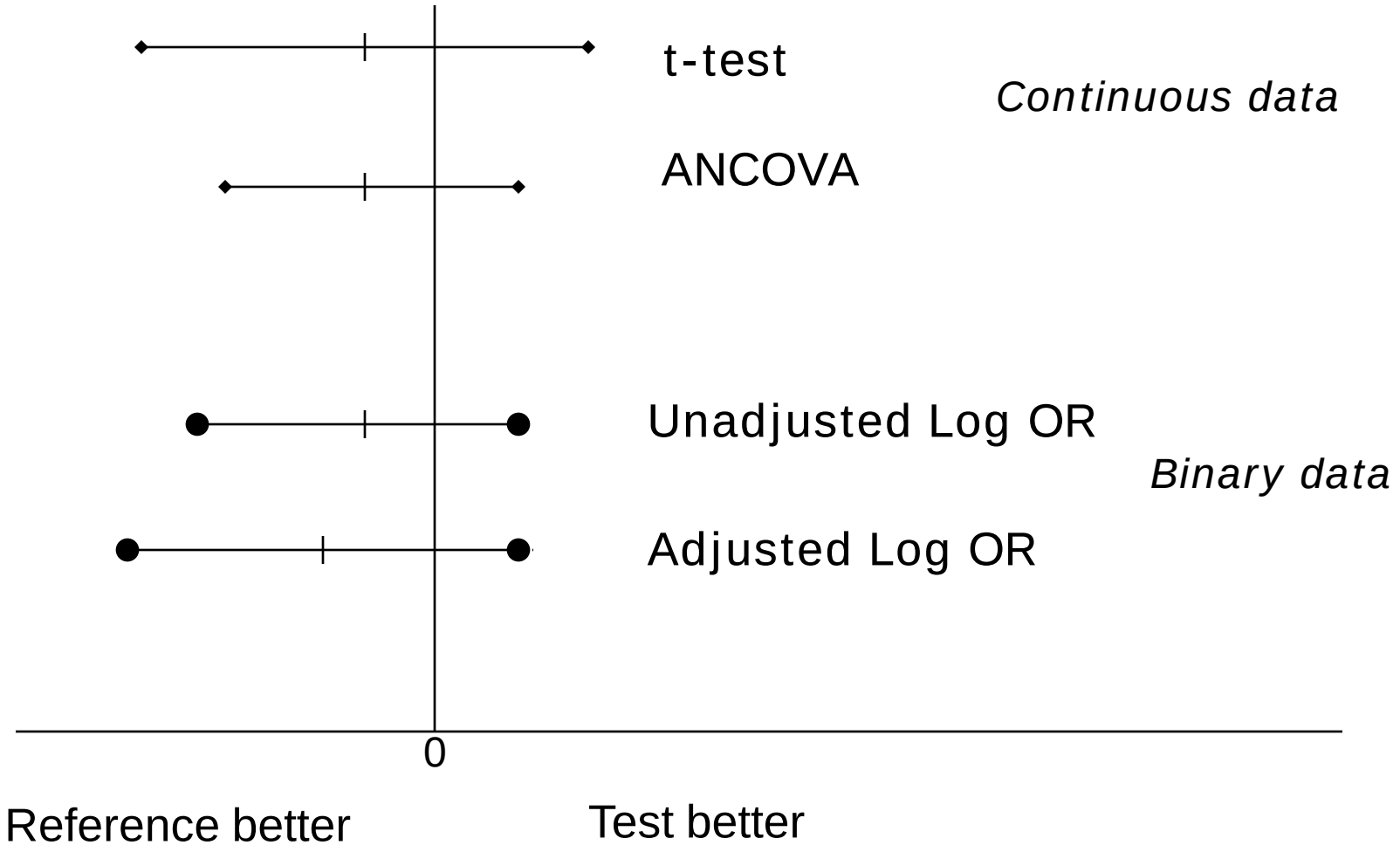
Stokes MA, Davis CS, Koch GG. *Categorical data analysis using the SAS system (2000)*

Covariate adjustment

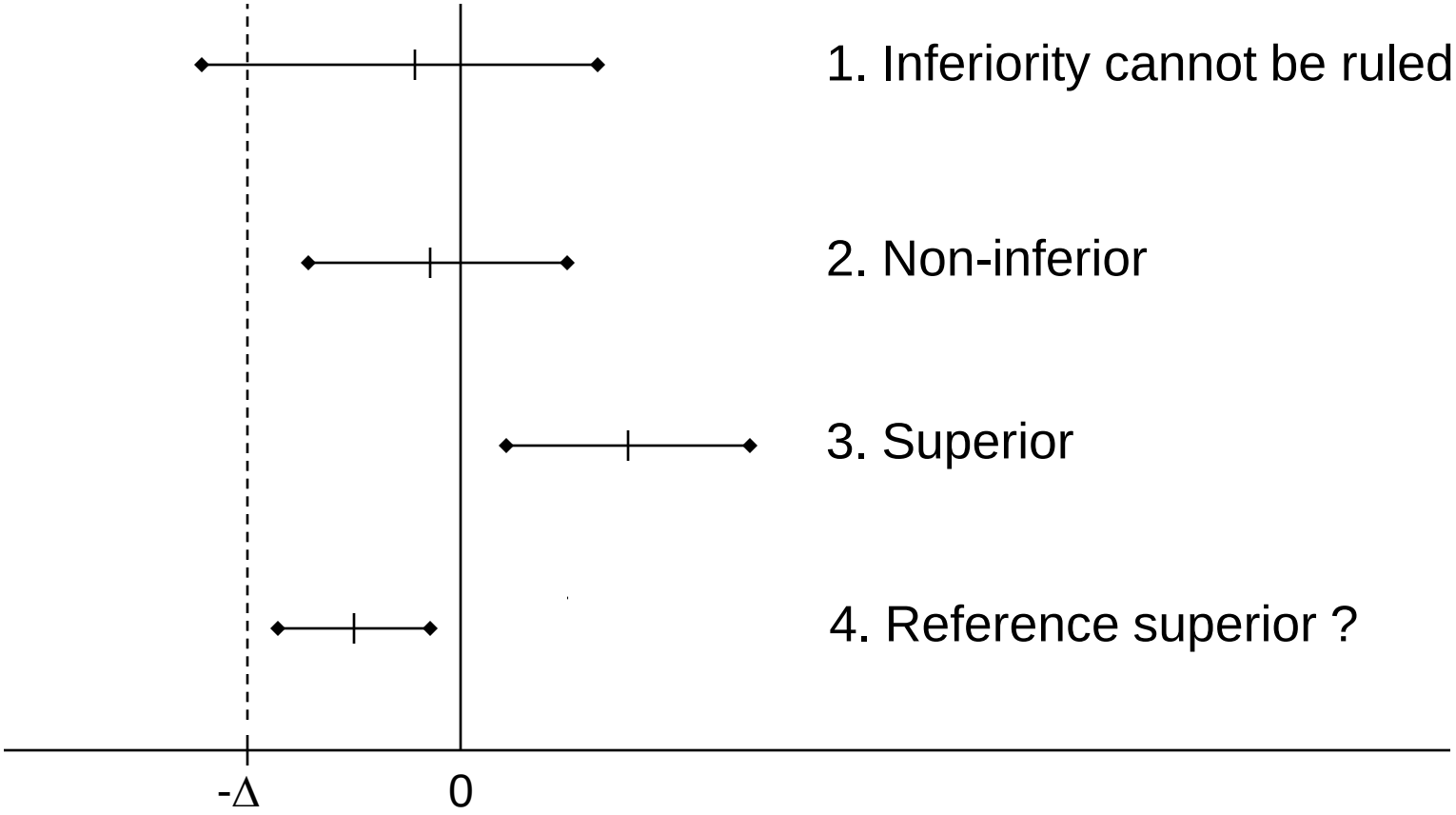
- If a factor exists which independently affects outcome then **excluding** this factor from the logistic model leads to:
  - underestimation of a non unity treatment difference
  - an increase in precision for the estimated treatment difference
- Combined effect of including a factor is an increase in efficiency - strategy of covariate adjustment justified for superiority - but what about non-inferiority

- Robinson *et al* (1991, 1998)

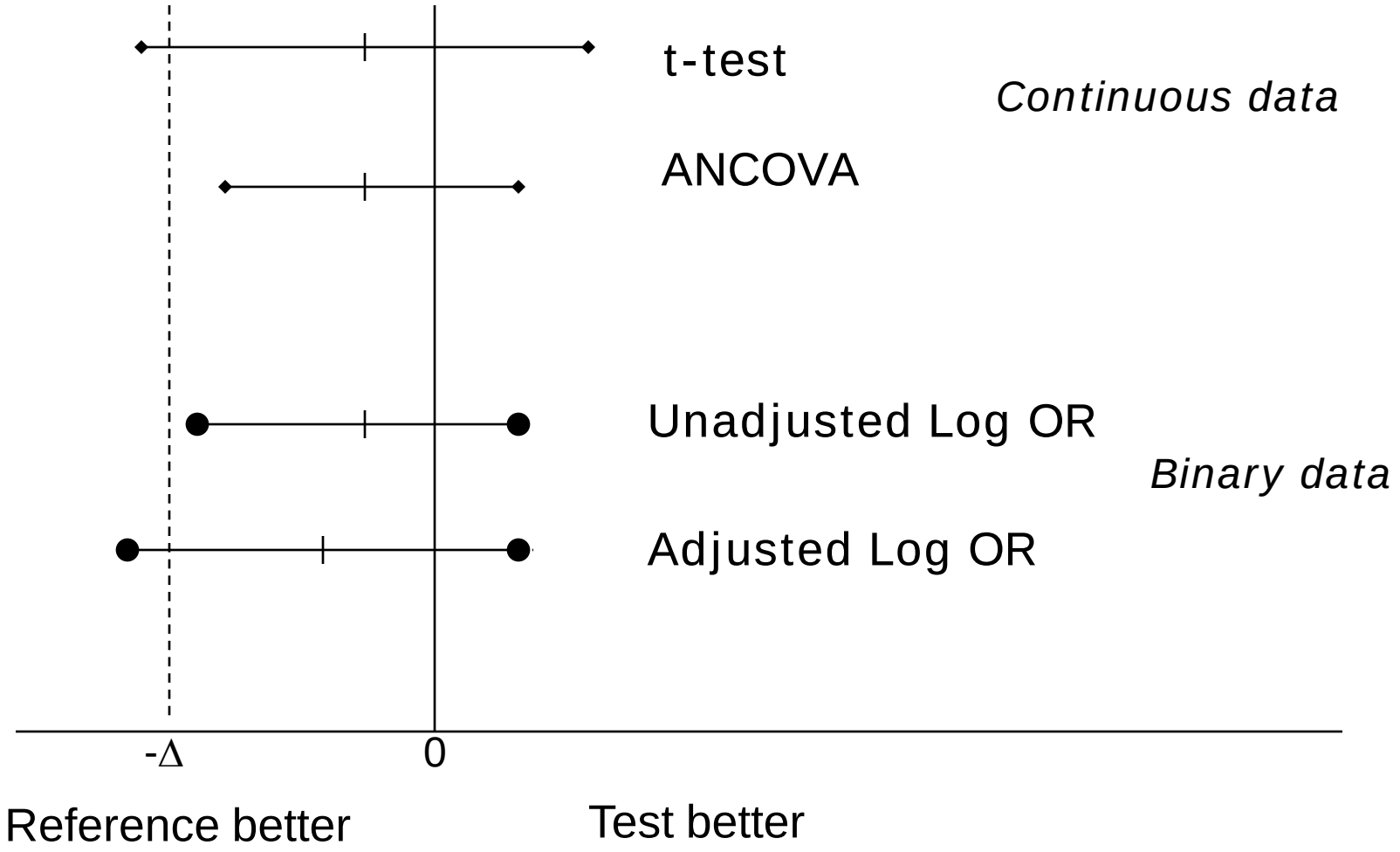
Impact of factor adjustment on Log OR & SE



Non-inferiority



Impact of factor adjustment on Log OR & SE: Effect on NI trials



Simulation (logistic model)

ψ_F	Model	% non inferior ψ_T =0.538	ψ_T =1
1	T	2.74	82.10
	F+T	2.58	81.94
2	T	2.80	81.14
	F+T	2.20	79.98
3	T	3.62	82.78
	F+T	2.52	79.86
4	**T**	**5.02**	**82.32**
	F+T	**2.78**	**76.52**

N=5000 simulations with sample size of 175 per treatment to show non-inferiority within 15% of a Reference percentage of 50% (one sided type I error of 2.5%, 80% power). F is a two level factor (Garrett, 2003)

IMPACT: NI, logistic & covariate adjustment

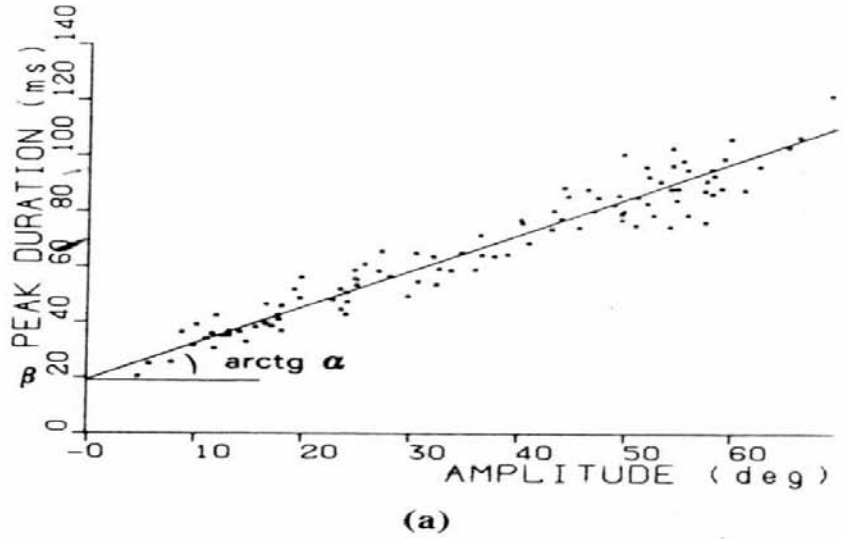
- When the factor effect is large then the type I error is approximately doubled if the factor is excluded
  - type I error: conclude non-inferiority when really inferior
- However unadjusted model has greater power - since when there is no treatment difference, covariate adjustment increases the SE.
  - Power: conclude non-inferiority when really non-inferior

Simpson's Paradox
Historical comparison of kidney stone removal (Charig et al, 1986)

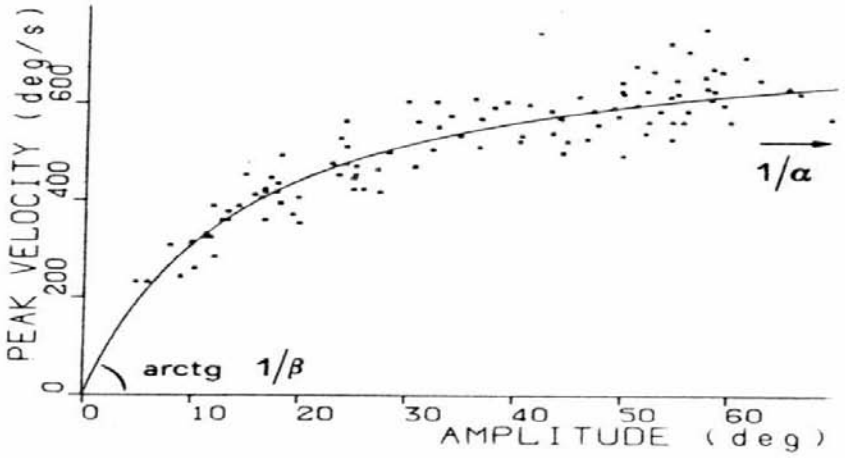
	<2 cm Diff= +6% OR=2.1	>=2 cm Diff= +4% OR=1.2	Total Diff= -5% OR=0.7
OS, 1972-80	**81/87=93%** odds =13.5	**192/263=73%** odds=2.7	**273/350=78%** odds=3.5
PN, 1980-85	**234/270=87%** odds=6.5	**55/80= 69%** odds=2.2	**289/350=83%** odds=4.7

Percutaneous nephrolithotomy (PN) vs. open surgery (OS) by stone diameter

Actual RCT example: Gaucher's disease: SEMV parameter estimation



(a)



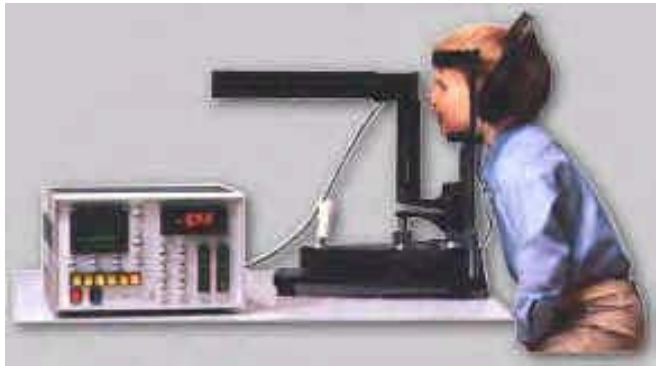
(b)

↓ slope (α) = improvement

↓ intercept (β) = improvement

↑ $1/\alpha$ = improvement

↑ $1/\beta$ = improvement



Simpson's Paradox: Gaucher's disease RCT

	Centre 1 Diff **-0.869**	Centre 2 Diff **-0.284**	Total Diff **0.190**
Test	**0.554** (14)	**-0.534** (4)	**0.312** (18)
Control	**1.423** (2)	**-0.250** (7)	**0.122** (9)

- 30 patients enrolled in RCT with 2:1 randomisation (active : placebo)
- Primary endpoint saccadic eye movement (quantitative endpoint)
- Mean change in slope α from BL to M12
- Study not stratified by centre as open-label design (avoid selection bias) led to some imbalance – exaggerated by:
  1. One patient refused Test, and followed as if Control
  2. Two patients in Test group and one in Control group with no saccade data

RSS First in man WP (TGN1412 study)

- A monoclonal antibody (rheumatoid arthritis, leukaemia etc)
- First-in-man study (13MAR06) conducted by Parexel on behalf of TeGenero
- In first cohort 8 volunteers
  - Six allocated TGN1412 and two allocated placebo
- All six given TGN1412 suffered a cytokine storm and by that evening had been admitted to intensive care

worst affected, suffering heart, liver and kidney failure, pneumonia, and septicaemia.

Study Outcome in a Conventional Format

		Adverse Reaction?		
		No	**Yes**	**Total**
Treatment	Placebo	2	0	2
	TGN 1412	0	6	6
	Total	2	6	8

Table 1. Fourfold summary showing healthy volunteers cross-classified by treatment and outcome.

Step	TGN1412		Placebo
	Dose Mg/Kg bodyweight	Number of Subjects	Number of Subjects
1	0.1	6	2
2	0.5	6	2
3	2.0	6	2
4	5.0	6	2

Table 1: Design given in the protocol

Questions

1. Why 6 + 2
2. Why four doses?
3. Why those dose increases?
4. Why simultaneous treatment in cohorts?

Adverse event classification (Strom, 1995):

1. Type A: predictable; dose related; less severe, extension of pharmacological effect
2. Type B: unpredictable, severe not related to dose (potentially hypersensitivity, immunological reactions)

To pool across cohorts or not?

- ‘Yes’, according to the protocol
- But then misleading to describe the trial as double-blind

- Bias variance trade-off
  - The proposed analysis would not eliminate the biases blinding is designed to eliminate
- Also what about the analysis at the end of each dose step to guide dose-escalation
- How would this permit pooling of placebo subjects

Design	Number of subjects					Variance of differences between doses, and between placebo and each dose, if							
						a cohort effect is fitted				it is known that there is no cohort effect			
1	Dose	0	1	2	3		1	2	3		1	2	3
	Cohort 1	2	6	0	0	0	0.67	0.67	0.67	0	0.33	0.33	0.33
	Cohort 2	2	0	6	0	1		1.33	1.33	1		0.33	0.33
	Cohort 3	2	0	0	6	2			1.33	2			0.33
2	Dose	0	1	2	3		1	2	3		1	2	3
	Cohort 1	4	4	0	0	0	0.50	0.50	0.50	0	0.33	0.33	0.33
	Cohort 2	4	0	4	0	1		1.00	1.00	1		0.50	0.50
	Cohort 3	4	0	0	4	2			1.00	2			0.50
3	Dose	0	1	2	3		1	2	3		1	2	3
	Cohort 1	4	4	0	0	0	0.29	0.40	0.65	0	0.29	0.31	0.39
	Cohort 2	2	2	4	0	1		0.40	0.65	1		0.31	0.39
	Cohort 3	1	1	2	4	2			0.58	2			0.42

QUINTILES

Rosemary Bailey's work in Senn (2007)

Bailey recommends designs that follow:

- Halving principle, in which no treatment is allocated to more than half of the subjects
- Diversity principle, where as many different treatments as possible are applied in each cohort
- Extra cohort principle, which means that there should be one more cohort than doses.

Contrast calculations

- In cohort i (n=6 dose vs n=2 PB):
- Var (dose i – PB) = Var (dose i) + Var (PB)
- $(1/6 + 1/2)\ \sigma^2 = 2/3\ \sigma^2$

- Ignoring cohort i (n=6 dose vs n=6 PB):

- $(1/6 + 1/6)\ \sigma^2 = 2/6\ \sigma^2$

Current areas of statistics research in the industry

Adaptive designs

Missing data

Non-inferiority and relative effectiveness

Bayesian

Analytics/data mining

References

- Charig CR, Webb DR, Payne S.R, Wickham OE. Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy. *British Medical Journal* 1986; **292:**879-882.
- Garrett AD. Therapeutic equivalence: fallacies and falsification. *Statistics in Medicine* 2003; **22**: 741-762.
- Koch G, Carr G et al. In: *Statistical methodology in the pharmaceutical sciences* (Ed. Berry DA) . Marcel Dekker, New York, 1990.
- Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review* 1991; **58**(2):227-240..
- Robinson LD, Dorroh JR, Lein D, Tiku ML. The effects of covariate adjustment in generalized linear models. *Communications in Statistics. – Theory and Methods* 1998; **27**:1653-1675.
- Simpson EH. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B* 1951; **13**:238-241.
- Smith C, Burley C, Ireson M. Clinical trials of antibacterial agents: a practical guide to design and analysis. *Journal of Antimicrobial Chemotherapy* 1998; **41**:467-480.
- Senn S, Julious S. Measurement in clinical trials: A neglected issue for statisticians? *Statistics in Medicine* 2009; **28**: 3189-3209.
- Senn *et al*. Working Party on Statistical issues in First-in-Man Studies. Statistical Issues in first-in-man studies. *JRSS Series A* 2007; **170**: 517-579.
- Stokes MA, Davis CS, Koch GG. *Categorical data analysis using the SAS system (2nd edition).* SAS Institute Inc.: Cary, NC, 2000.