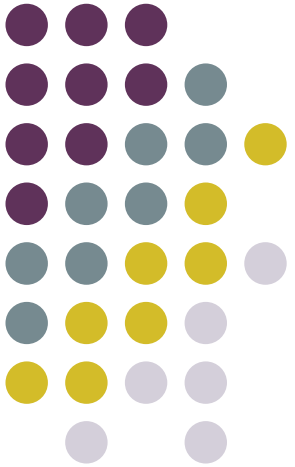
# Genetics and the design of animal experiments

Michael FW Festing, Ph.D., D.Sc., CStat.

michaelfesting@aol.com



# Some well known examples of medical advances involving animal experimentation

- Anti-rabies vaccine developed c1885 using dogs and rabbits (Pasteur)
- Salvarsan for treating syphilis c1911 following screening of >600 arsenic compounds in rabbits (Ehrlich)
- Penicillin isolated (c1940 by Fleming, in-vitro test) and tested for toxicity and efficacy using mice (Chain & Florey)
- Blood transfusion 1900-1916
- HIV and anti-retroviral therapy 1986-1996
- Organ transplants from 1950
- Parkinson's disease and deep brain stimulation (recent)
- Meningitis vaccination 1992
- All modern drugs

# Principles of Humane Experimental Technique

Russell and Burch (1959)

**Replacement:**

Where possible non-sentient alternatives should be used

**Refinement:**

Pain and distress should be minimised.

**Reduction**:

The minimum number of animals should be used consistent with achieving the objectives of the study.

A cost benefit analysis is also necessary. Is the experiment worth doing? (often left to the funding organisations)

# 93% of research involves mice, fish and rats

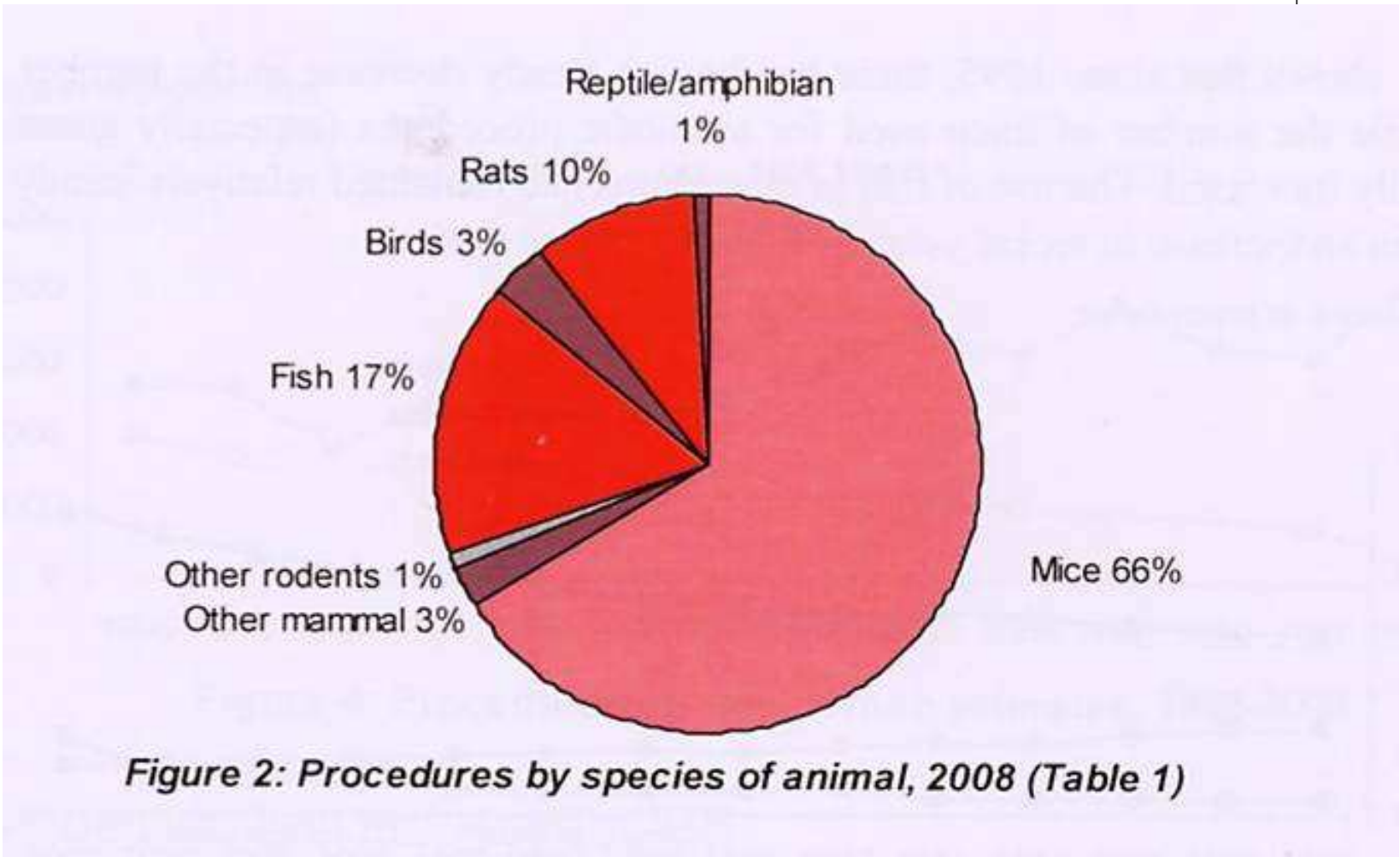



*Figure 2: Procedures by species of animal, 2008 (Table 1)*

# Use of vertebrate animals in biomedical research and testing in the UK, 2008

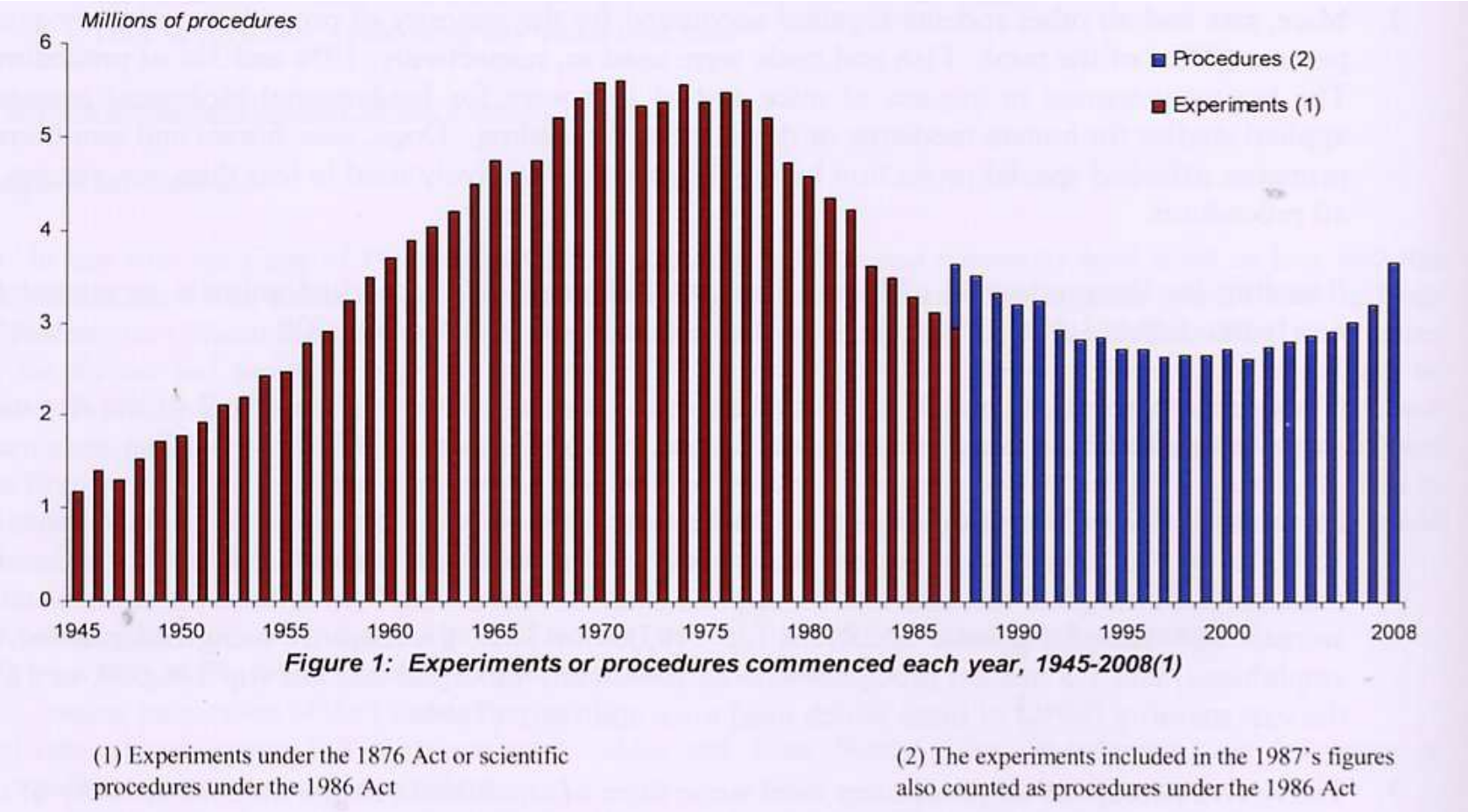



*Figure 1: Experiments or procedures commenced each year, 1945-2008(1)*

(1) Experiments under the 1876 Act or scientific procedures under the 1986 Act

(2) The experiments included in the 1987's figures also counted as procedures under the 1986 Act

Statistics of scientific procedures on living animals 2008, London Stationary Office

# Genetically modified strains

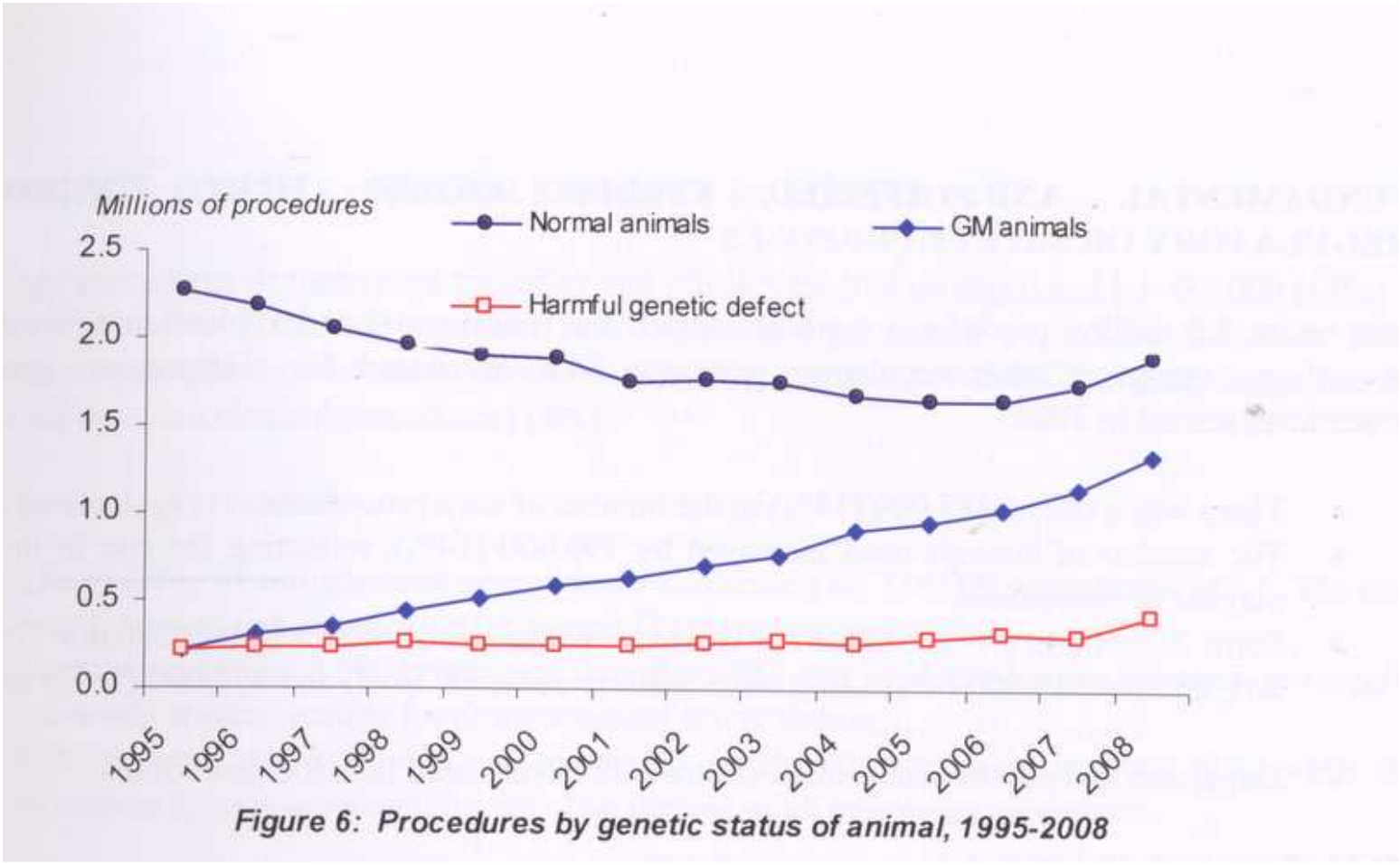



*Figure 6: Procedures by genetic status of animal, 1995-2008*

# Genetically modified strains

- Usually on an inbred genetic background, particularly C57BL/6
- Transgenic animals. Incorporation of foreign DNA
  - E.g. the Oncomouse, Big Blue, Immortomouse (SV40 T antigen), gfp etc.
- Knockouts
  - Already >4000 targeted mutants
  - And >7000 unique genes trapped in ES cells
- Knockins
  - Targeted insertion, e.g. “humanised” mice with human Cyp genes

# Technical advances leading to the use of genetically modified mice

- Inbred strains  (Little et al since 1909)
- Embryo freezing (Whittingham c1970)
- Embryonic stem cells (Martin Evans 1981)
  - Used the 129 strain of mice
- DNA handling/sequencing
  - Human complete DNA sequence 2004
  - Mouse final DNA sequence, C57BL/6J, 2009
- Methods of mutating mouse genes
  - Gene trapping in ES cells  (c 1989)
  - Homologous recombination (Capecchi, Evans, Smithies, 1989)
  - ENU mutagenesis

# Build 36. Complete DNA sequence of the C57BL/6J mouse

- 20,210 protein coding genes, compared with 19,042 in humans
- 15,187 human/mouse orthologs
  - 80% of human genes have a counterpart in mice
  - Most mouse-specific genes are to do with reproduction
- Draft sequence of another 16 inbred strains has been published

# International knockout mouse consortium (IKMC)

- Plan to knock out every gene in the mouse
  - ~$150 million
  - Will use C57BL/6 ES cells
- Use gene trapping in ES cells (“shotgun” approach)
- Then use targeted mutagenesis
- Use of sophisticated targeting vectors
  - Specific tissues
  - Specific times
- Mutant ES cells will be made available to investigators free of charge

# Animals are used as models

- Animals are usually used as models of humans or domestic animals
  - There are an infinite number of models to choose from
  - Our experiments only tell us what happens to the model, not to humans.
- We have substantial control of the animals and their environment
  - experiments can be relatively small yet still powerful
  - they can be quite complex
- Need for a multi-disciplinary approach
  - Veterinary science, species biology, animal husbandry, genetics, statistics, surgery etc.

# Infinite number of animal models

- >12 species
- Mouse 66% of animals used
  - >400 inbred strains
  - >500? mutants, expression may depend on genetic background
  - >10,000? genetic modifications
    - Project to “knock out” all 22,000 mouse genes
  - 8.5 million SNPs in 15 inbred strains
  - Many surgical and drug-induced models
  - >100 models of breast cancer in mice

# A good model

- Biologically relevant
  - Sensitive
  - Repeatable
  - Powerful (good discrimination)
- Experimentally convenient
  - Good knowledge base
  - Economical
  - Easily available
  - Ethically acceptable

# Is animal research being done well?

# Survey of statistical quality of published papers

McCance, 1995 Aust. Vet. Journal. 72:322

| | |
|---|---|
| Number of papers studied | 133 |
| | % |
| • Require statistical revision | 61 |
| • Serious errors | 5 |
| • Deficiencies in design | 30 |
| ◦ (randomisation, size, heterogeneity, bias) | |
| • Deficiencies in statistical analysis | 45 |
| ◦ (sub-optimal methods, errors calculation) | |
| • Deficiencies in presentation | 33 |
| ◦ (omission of data, stats. inappropriate) | |

# Survey of 271 papers. Results published in PLoS One

Of the papers studied:

- 5% did not clearly state the purpose of the study
- 6% did not indicate how many separate experiments were done
- 13% did not identify the experimental unit
- 26% failed to state the sex of the animals
- 24% reported neither age not weight of animals
- 4% did not mention the number of animals used
- 0% justified the sample sizes used
- 35% which reported numbers used these differed in the materials and methods and the results sections
- etc.

Kilkenny et al (2009), PLoS One Vol. 4, e7824

# The ARRIVE guidelines

- (Animals in Research: Reporting In Vivo Experiments)
- Check list of 20 items to be considered when reporting research results
- Based on the CONSORT guidelines

Kilkenny et al (2010). PLoS Biology | www.plosbiology.org. Volume 8. Issue 6 e1000412

# A case study

## Do toxicologists use the wrong model?

Current model is the **outbred** Sprague-Dawley rat or CD-1 mouse.

An alternative is the Multi-Strain Assay (MSA), a small collection of **inbred** strains using the same total number of animals (Russell and Burch 1959)

# Toxicity testing

Acute toxicity

28-day & 90-day repeat dose toxicity test

- Control, low, medium, high dose
- 10 animals of each sex/dose (80 animals total)
- Measure haematology, clinical chemistry, organ weights, histopathology, etc.

2-year carcinogenesis study

Reproductive toxicity

Other tests

# Current assays lack statistical power

"The traditional tools used to assess product safety -- animal toxicology and outcomes from human studies -- have changed little over many decades and have largely not benefited from recent gains in scientific knowledge. The inability to better assess and predict product safety leads to failures during clinical development and, occasionally, after marketing."

*FDA Critical Path White Paper, 2004*

# 1099 Investigative new drugs

96% attrition rate

27% due to toxicity

Caldwell et al,. Curr Top Med Chem 2001; 1(5):353-366.

# Suggested use of inbred strains in toxicity testing

"Toxicity testing, as usual.., is the scene of some confused thought, which may be delaying the exploitation of statistical methods.

……..

We have not infrequently heard the opinion expressed that…. in toxicity tests you need a thoroughly heterogeneous mass of animals, and plenty of them.

"The fallacy consists in supposing that in order to obtain a wide inductive base a heterogeneous stock should be used……

The proper procedure is, of course to use several different homogeneous samples, by using a plurality of pure lines (or preferably F1 crossbreds)…for otherwise the experimenter deprives himself of the possibility of making a relatively precise estimate of the error (Fisher 1942).""

*(Russell and Burch 1959)*

# Inbred strains and outbred stocks of mice and rats

Inbred (isogenic) strain

- Isogenic (animals identical)
- Homozygous, breed true
- Phenotypically uniform
- Defined (quality control)
- Genetically stable

Like immortal clones of genetically identical individuals.

Outbred stocks

- Each individual different
- Do not breed true
- Phenotypically variable
- Not defined (no QC)
- Genetic drift can be rapid

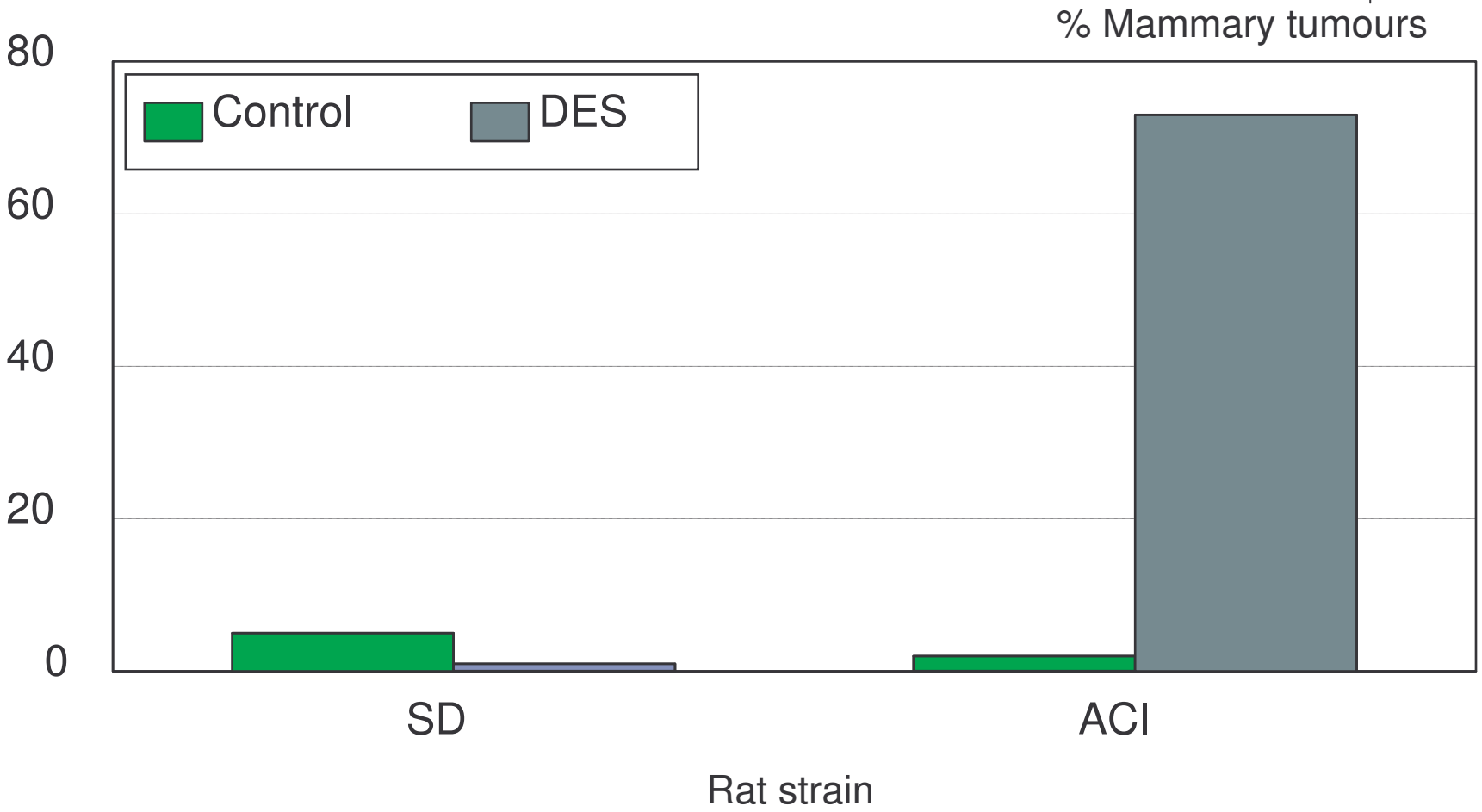
Genetically heteogeneous like a human population.

Almost all toxicity testing is currently done using outbred stocks such as Sprague-Dawley and Wistar rats and CD-1 and Swiss mice

# Papers suggesting the use of a multi-strain assay in toxicity testing (Festing)

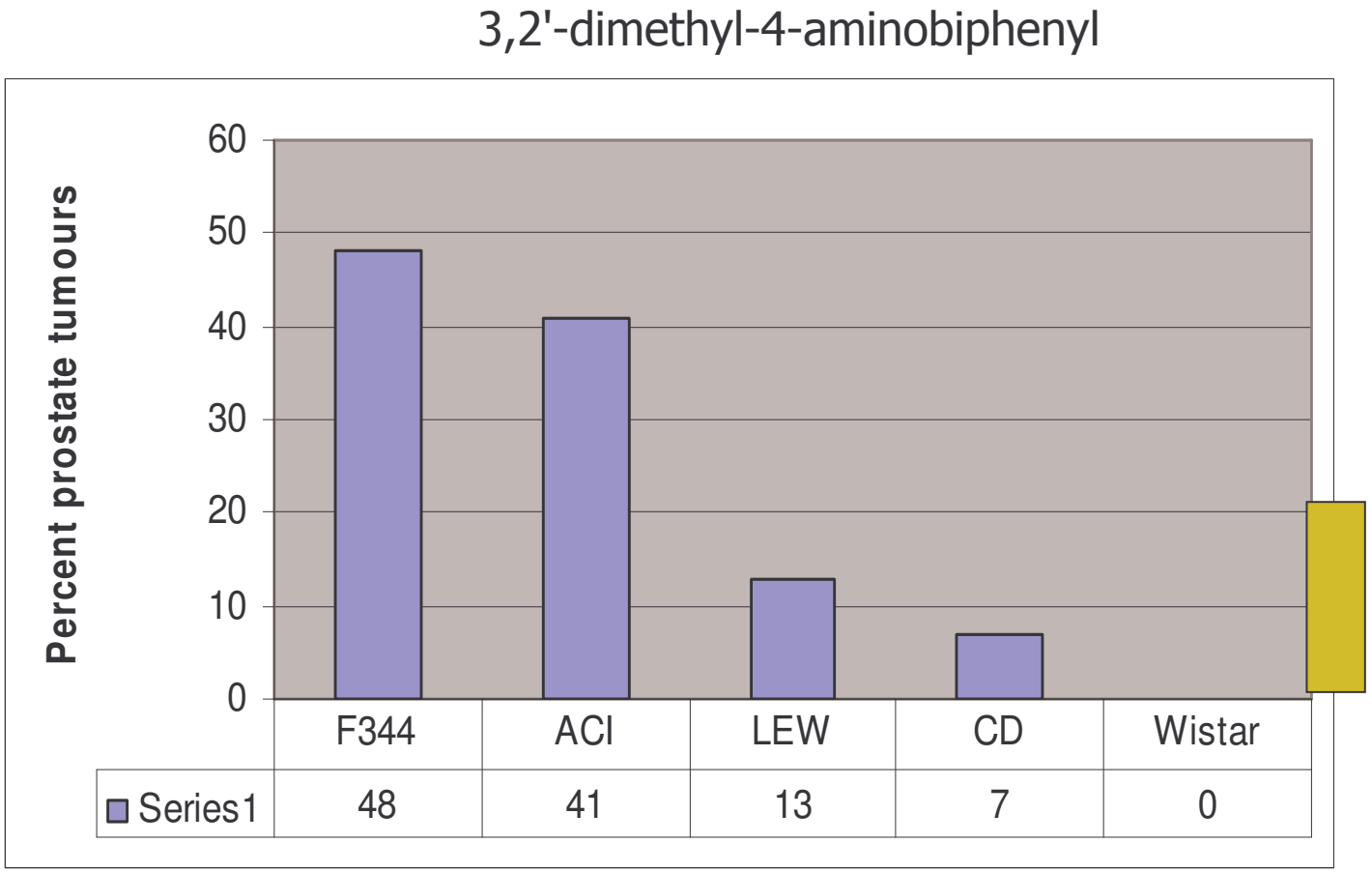
1975 Food and Cosmetics Toxicology
1979 J. of Toxicology and Environmental Health
1980 Acta Zoologica et Pathologica Antvipensia
1986 Archives of Toxicology, Supplement
1987 CRC Critical Reviews in Toxicology
1990 Toxicology and Applied Pharmacology
1991 Experientia
1993 J. of Experimental Animal Science
1995 Environmental Health Perspectives
1996 J. of the Royal Statistical Society Series B-Methodological
1997 Nature
1997 Comparative Haematology International
1999 Neurobiology of Aging
2001 Food and Chemical Toxicology
2005 Nature Genetics
2010 Toxicol.Pathol.
2010 Methods Mol.Biol. 2010;602:1-21.

# The results of a toxicity study may depend entirely on the strain of animals used. Carcinogenic effects of DES




Shellabarger et al J.Nat.Can.Inst. 61:1505, 1978. Results averaged over 4 irradiation levels, N>100 in each group.

# Power increased by using > one strain/stock




10 rats of each or 5 strains 21%

Shirai et al (1990) Carcinogenesis 11:793

# Using more than one strain/ stock in a toxicity screen

- If strains don't differ, then nothing is lost
- Where strains differ a multi-strain strategy is most powerful
- The best strategy is to use as many strains as possible
- The only situation where the use of a single strain is best is if the chosen strain is the most sensitive (a rare situation)

Felton,R.P. & Gaylor,D.W. Multistrain experiments for screening toxic substances. *Journal of Toxicology and Environmental Health* **26**, 399-411 (1989).

# Genetic heterogeneity leads to low powered experiments or large sample sizes

Sample sizes needed to detect a 4 minute difference in hexobarbital sleeping time between a treated and control group

| Strain | Mean* | SD | Sig/noise | Group size** | Power*** |
|---|---|---|---|---|---|
| A/N | 48 | 4 | 1.0 | 23 | 86 |
| BALB/c | 41 | 2 | 2.0 | 7 | 99 |
| C57BL/HeN | 33 | 3 | 1.3 | 13 | 98 |
| C3HB/He | 22 | 3 | 1.3 | 13 | 98 |
| SWR/HeN | 18 | 4 | 1.0 | 23 | 86 |
| CFW | 48 | 12 | 0.3 | 191 | 17 |
| Swiss | 43 | 15 | 0.2 | 297 | 13 |

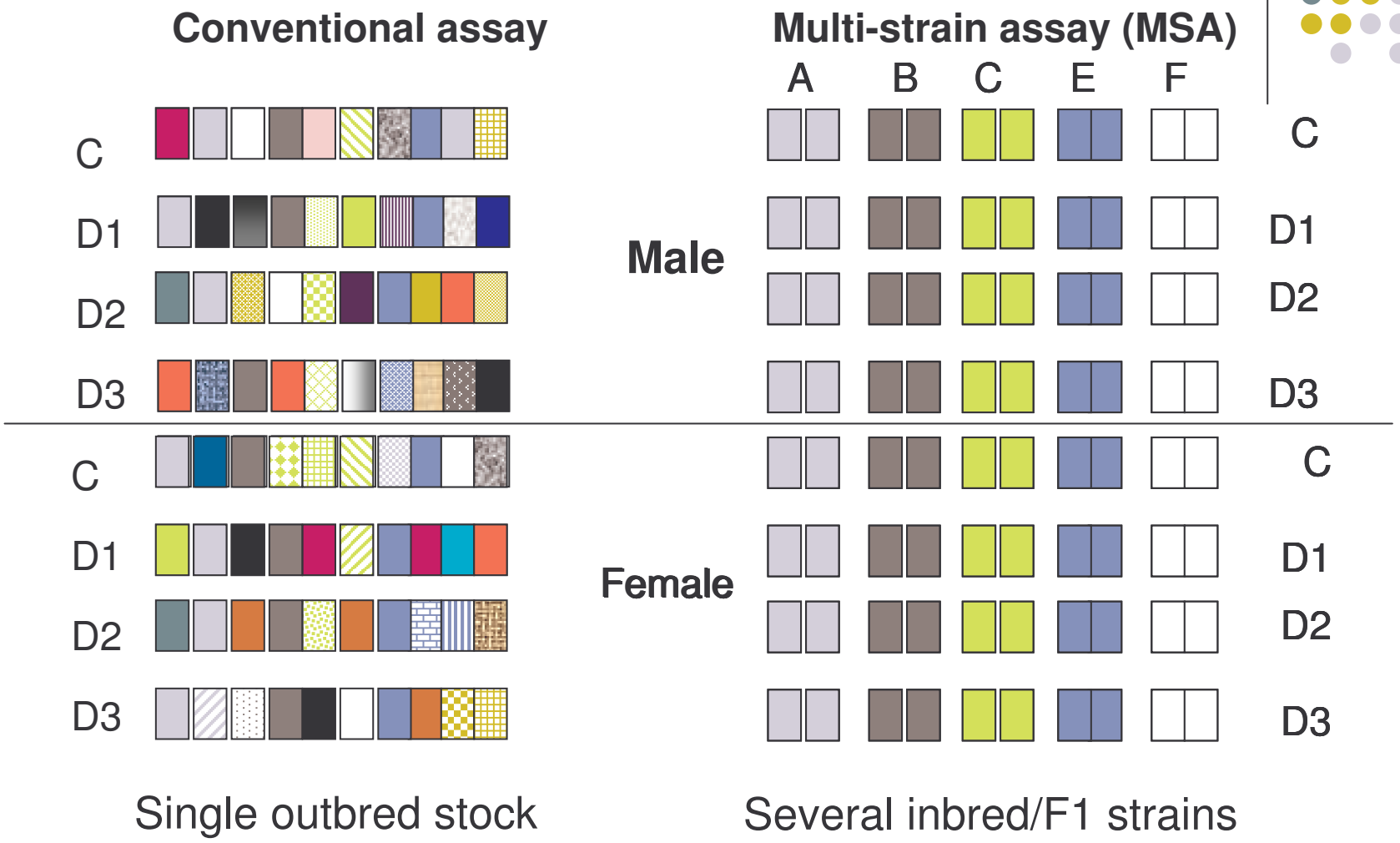
*N= 25-47

** Assumes a 2-sided t-test with $\alpha$=0.05, power = 90%

*** Assumes a fixed sample size of 20 mice/group

Data from Jay 1955 Proc Soc. Exp Biol Med 90:378

# The Multi-Strain 28-day assay: two possible designs




Festing MF. Inbred strains should replace outbred stocks in toxicology, safety testing, and drug development. *Toxicol.Pathol.* 2010; **38:** 681-690.

# Example: Only data on White blood cell counts at top dose level shown

Dose of chloramphenicol (mg/kg) given by gavage, 7 days

| | 0 | 500 | 1000 | 1500 | 2000 | 2500 |
|---|---|---|---|---|---|---|
| Outbred | | | | | | |
| CD-1 | 8 | 8 | 8 | 8 | 8 | 8 |
| MSA (Inbreds) | | | | | | |
| CBA | 2 | 2 | 2 | 2 | 2 | 2 |
| C3H | 2 | 2 | 2 | 2 | 2 | 2 |
| BALB/c | 2 | 2 | 2 | 2 | 2 | 2 |
| C57BL | 2 | 2 | 2 | 2 | 2 | 2 |

Festing,M.F.W.,et. al. (2001) Strain differences in haematological response to chloramphenicol succinate in mice: implications for toxicological research. *Food and Chemical Toxicology*, **39**, 375-383.

# White blood cell counts following treatment with chloramphenicol at 2500mg/kg

White blood cell counts

| Strain | Control | Treated |
|---|---|---|
| CBA | 1.90 | 0.40 |
| CBA | 2.60 | 0.20 |
| C3H | 2.10 | 0.40 |
| C3H | 2.20 | 0.40 |
| BALB/c | 1.60 | 1.30 |
| BALB/c | 0.50 | 1.40 |
| C57BL | 2.30 | 0.80 |
| C57BL | 2.20 | 1.10 |
| CD-1 | 3.00 | 1.90 |
| CD-1 | 1.70 | 1.90 |
| CD-1 | 1.50 | 3.50 |
| CD-1 | 2.00 | 1.20 |
| CD-1 | 3.80 | 2.30 |
| CD-1 | 0.90 | 1.00 |
| CD-1 | 2.60 | 1.30 |
| CD-1 | 2.30 | 1.60 |

Four inbred strains

One outbred stock

# WBC counts following chloramphenicol at 2500mg/kg

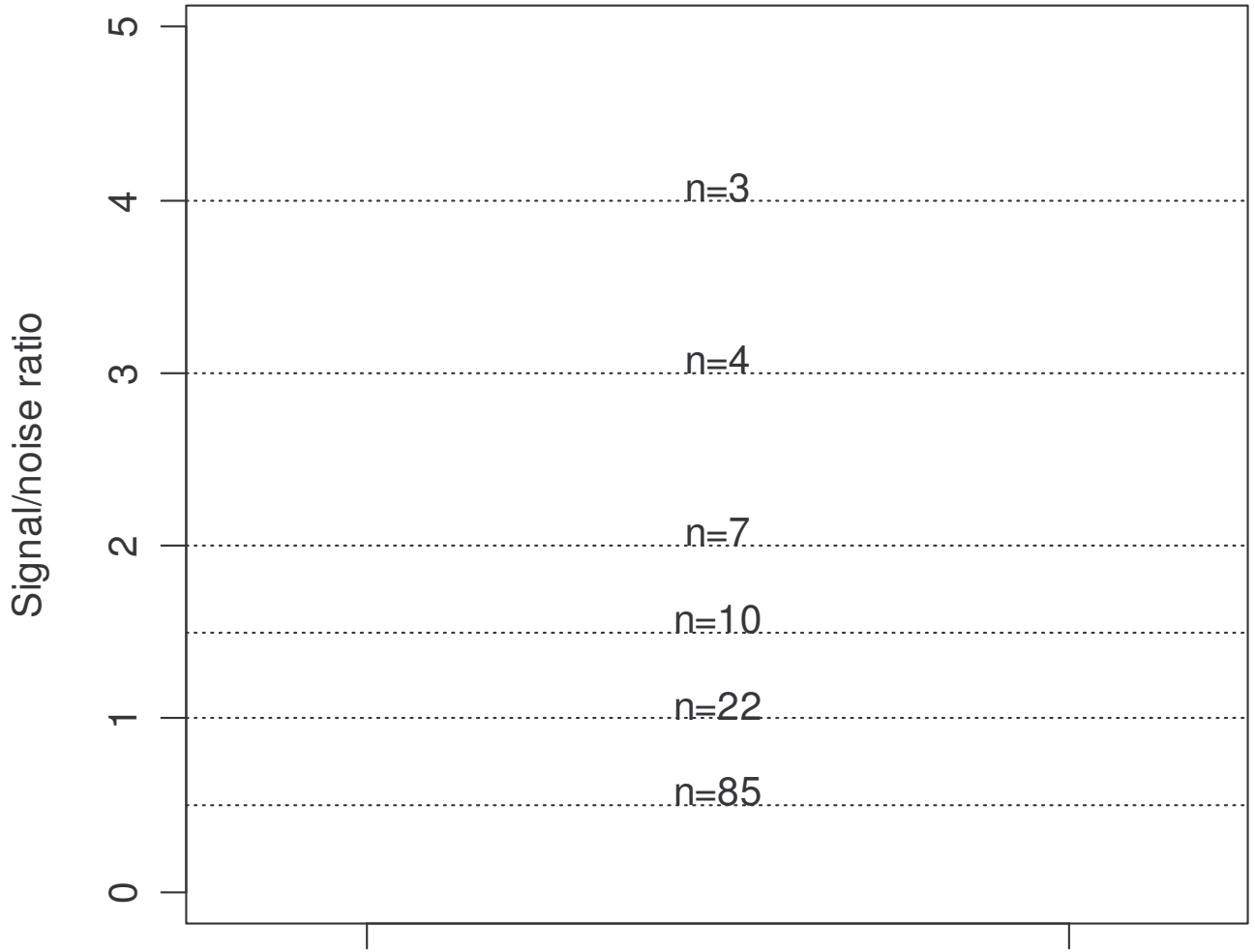
**White blood cell counts**

| Strain | N | 0 | 2500 | Signal (Difference) | Noise (SD) | Signal/noise | p |
|---|---|---|---|---|---|---|---|
| CD-1 | 16 | 2.23 | 1.83 | 0.40 | 0.86 | 0.47 | 0.38 |

| Strain | N | 0 | 2500 | Signal (Difference) | Noise (SD) | Signal/noise | p |
|---|---|---|---|---|---|---|---|
| CBA | 4 | 2.25 | 0.30 | 1.95 | 0.34 | 5.73 | |
| C3H | 4 | 2.15 | 0.40 | 1.85 | 0.34 | 5.44 | |
| BALB/c | 4 | 1.05 | 1.35 | (-0.30) | 0.34 | (-0.88) | |
| C57BL | 4 | 2.25 | 0.95 | 1.30 | 0.34 | 3.82 | |
| Mean | 16 | 1.93 | 1.20 | 0.73 | 0.34 | 2.15 | <0.001 |
| Dose * strain | | | | | | | <0.001 |

Signal/noise ratio= standardised effect size = (M1-M2)/SD

# Quantifying the advantage of the multi-strain assay in terms of sample size using a power analysis

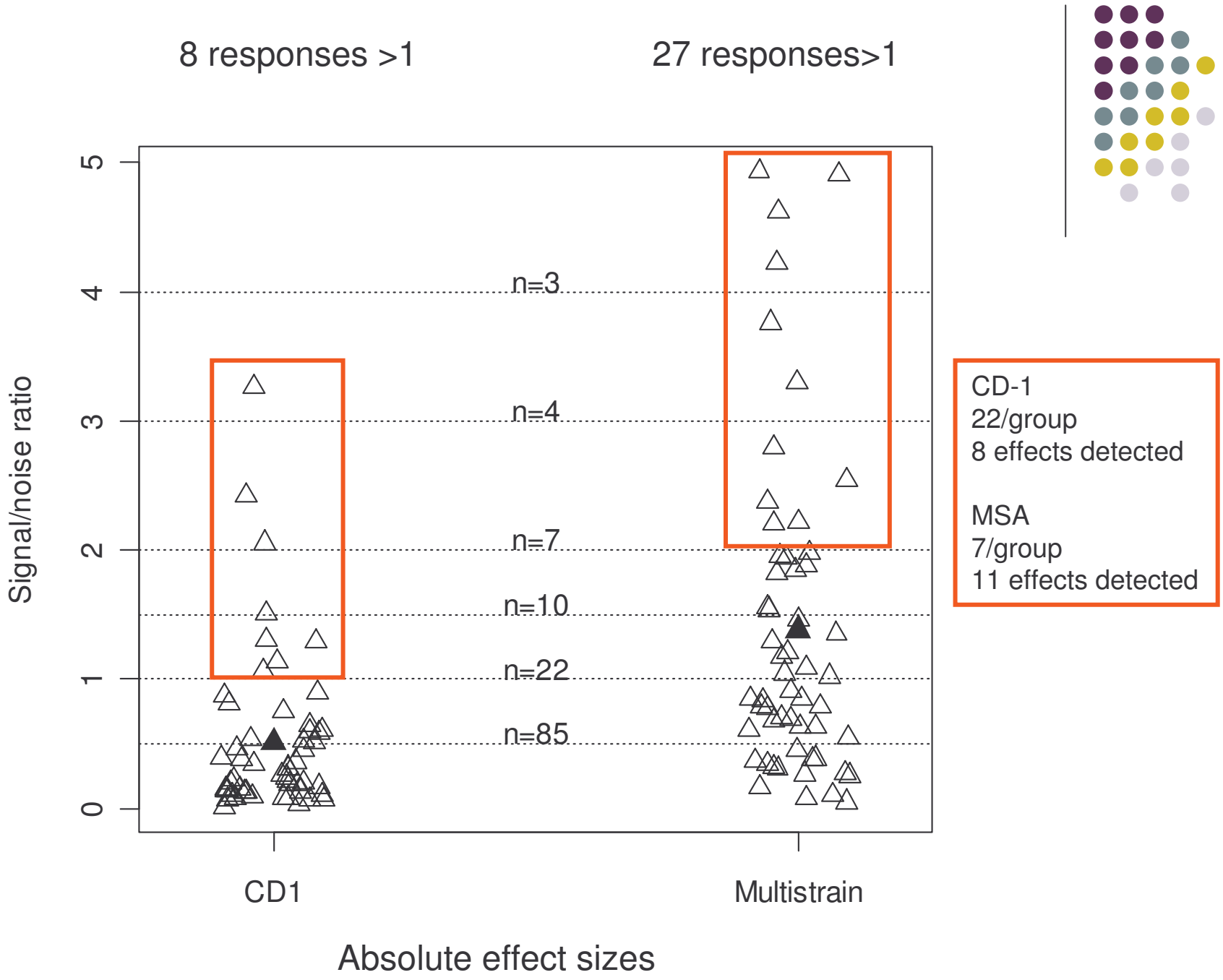



Assumes two-sample t-test:

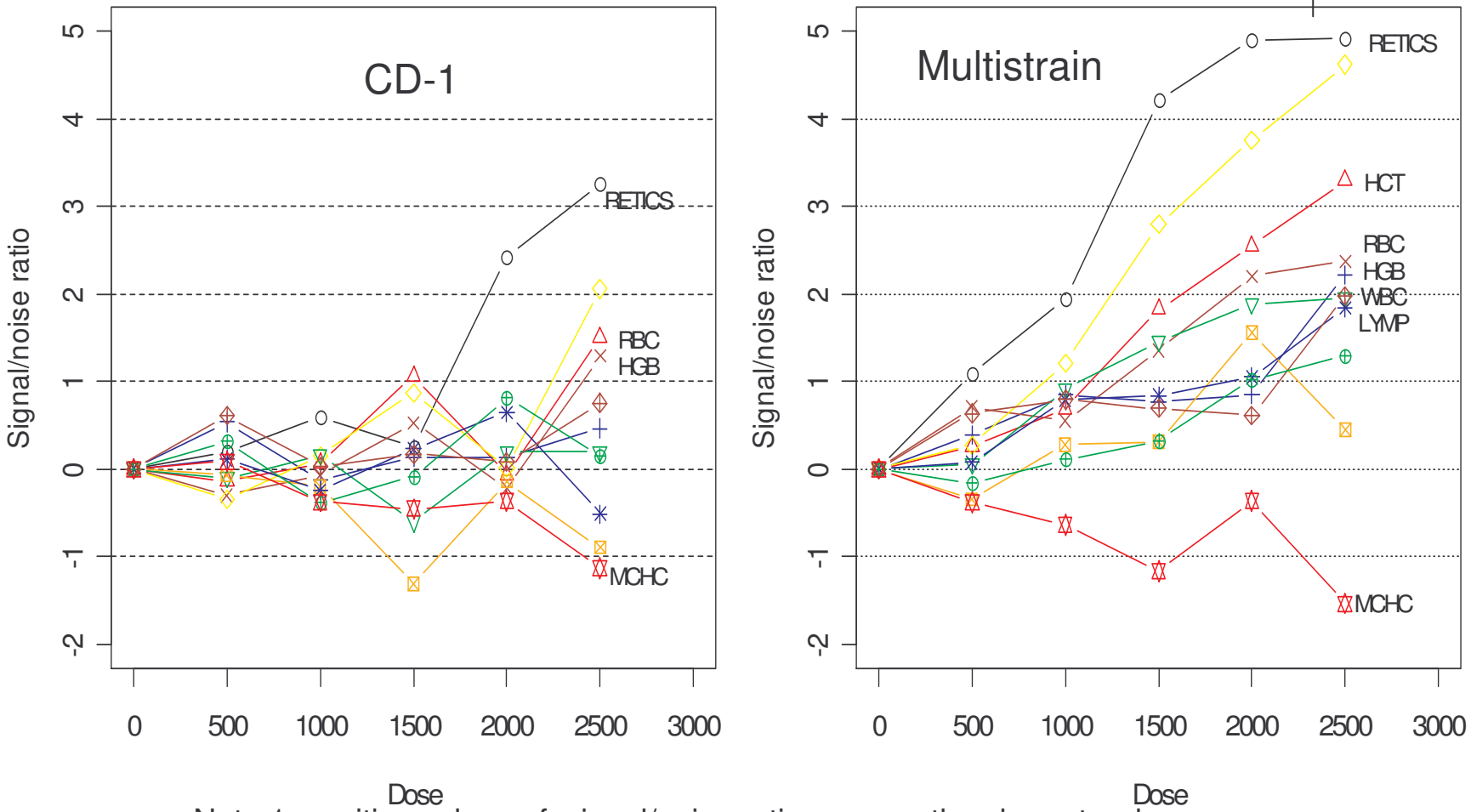
Power=90%

Significance level =0.05

Alternative 2-sided


8 responses >1
27 responses>1
Signal/noise ratio
0
1
2
3
4
5
n=3
n=4
n=7
n=10
n=22
n=85
CD1
Multistrain
Absolute effect sizes
CD-1
22/group
8 effects detected
MSA
7/group
11 effects detected

Response of 11 haematological parameters to chloramphenicol in CD-1 and the MSA. Note the better dose/responses in the MSA




Note 1. positive values of signal/noise ratios means the character decreases

2. Hematological characters are correlated

3. MSA is robust for missing observations

# Summary (chloramphenicol experiment)

- Multi-strain assay (MSA) was more powerful than single outbred stock
  - Equivalent results with 1/3$^{rd}$ the number of animals
  - But “payoff” should be taken in increased power not reduced sample size
- The MSA was robust to accidental losses up to about 25% of “deaths” (not shown)
- Strains differed in response (Genetics!)
  - CD-1 and BALB/c resistant but CBA, C3H and C57BL were susceptible
- MSA combined with “omics” could be even more powerful.

# Why have toxicologists not improved methods in >60 years?

- Conflict of interest
  - Mainly food and chemicals, but also pharmaceuticals until recently
- Fear of increased Type I (false positive) results
  - But decreasing Type II error does not automatically increase Type I error
  - Statistically significant results need to be explained
- Failure of leadership
  - FDA says “we don’t tell companies what data to present to us”
  - Companies “We provide whatever data the regulators require”.
- Bad science
  - Humans are outbred (high fidelity fallacy)
  - Poor understanding of statistics/experimental design (think the multi-strain design as having too small a group size)

# General conclusions

- Animal research has made an important contribution to the development of modern medicine
- Ethics is important with animals as well as humans, but informed consent not possible. Use “3Rs”.
- We have much more control over animals than we do over humans
  - Small sized experiments still powerful
  - Can inflict pain and death
  - Complicated experiments possible
- Important technical advances have led to the widespread use of genetically modified mice

# General conclusions (continued)

- Surveys suggest ample scope for improvement in the design & statistical analysis animal research
  - Failure to control genetic variability is widespread, particularly by users of rats
  - Methods of toxicity testing have hardly changed in more than 60 years and are extremely inaccurate
- Better training in experimental design & statistical analysis for investigators would save money and improve the quality of the science.