Some statistical issues in the design of experiments on animals or people

R. A. Bailey

r.a.bailey@qmul.ac.uk

Edinburgh, 18 August 2009

Biography

1965	9 months at the MRC's Air Pollution Research Unit at Barts
1965–68	Degree in Mathematics at Oxford
1968–69	VSO in Nigeria
1969–72	DPhil in Pure Mathematics at Oxford
1972–75	Open University
1976–78	Statistics post-doc at Edinburgh
1978–81	Open University
1981–90	Rothamsted Experimental Station (AFRC)
1991–94	Professor of Mathematical Sciences, Goldmsith's College, University of London
1994–	Professor of Statistics, Queen Mary, University of London

Three principles of experimental design

- Randomization
- Replication
- Control

Three principles of experimental design

- Randomization
  - Why do we randomize?
  - How do we randomize?
- Replication

- Control

Three principles of experimental design

- Randomization
  - Why do we randomize?
  - How do we randomize?
- Replication
  - Increased replication usually decreases variance.
  - Increased replication may increase variability.
  - Increased replication usually increases power.
  - Increased replication increases costs (monetary and human).
  - Beware of false replication.
- Control

Three principles of experimental design

- Randomization
  - Why do we randomize?
  - How do we randomize?
- Replication
  - Increased replication usually decreases variance.
  - Increased replication may increase variability.
  - Increased replication usually increases power.
  - Increased replication increases costs (monetary and human).
  - Beware of false replication.
- Control
  - Group the experimental units into blocks of alike units.
  - Concurrent comparison with "do nothing".
  - Concurrent comparison with at least one other treatment.

Why do we randomize?

It is to avoid

- systematic bias
  (for example, doing all the tests on treatment A in January
  then all the tests on treatment B in March)
- selection bias
  (for example, choosing the most healthy patients for the
  treatment that you are trying to prove is best)
- accidental bias
  (for example, using the first rats that the animal handler takes out
  of the cage for one treatment and the last rats for the other)
- cheating by the experimenter.

Lanarkshire milk experiment

Treatments: extra milk rations or not.
These should have been randomized to the children within each school.
The teachers decided to give the extra milk rations to those children who were most undernourished.

Doctor knows best

A consultant organized a trial of drugs to cure a serious disease. There were 3 treatments: the current standard drug X, which was a very strong antibiotic, and 2 new drugs. Several GPs agreed to participate in the trial. They were all sent the trial protocol, and asked to phone the consultant's secretary when they had a patient to be entered in the trial. The secretary had the randomization list, showing which drug to allocate to which patient in order as they entered the trial.

Doctor knows best

A consultant organized a trial of drugs to cure a serious disease. There were 3 treatments: the current standard drug X, which was a very strong antibiotic, and 2 new drugs. Several GPs agreed to participate in the trial. They were all sent the trial protocol, and asked to phone the consultant's secretary when they had a patient to be entered in the trial. The secretary had the randomization list, showing which drug to allocate to which patient in order as they entered the trial.

One day, a GP phoned and said that he had a suitable patient for the trial. The secretary asked several questions about age, weight etc., to check whether the patient was eligible and, if so, to determine the correct dose of the allocated drug. The secretary accepted the patient, allocated the next drug on the randomization list, which was X, worked out the dosage and told the GP that the patient should be given that dose of X. The GP said "My patient cannot take X, because it harms her." The secretary asked the consultant what to do.

Doctor knows best

A consultant organized a trial of drugs to cure a serious disease. There were 3 treatments: the current standard drug X, which was a very strong antibiotic, and 2 new drugs. Several GPs agreed to participate in the trial. They were all sent the trial protocol, and asked to phone the consultant's secretary when they had a patient to be entered in the trial. The secretary had the randomization list, showing which drug to allocate to which patient in order as they entered the trial.

One day, a GP phoned and said that he had a suitable patient for the trial. The secretary asked several questions about age, weight etc., to check whether the patient was eligible and, if so, to determine the correct dose of the allocated drug. The secretary accepted the patient, allocated the next drug on the randomization list, which was X, worked out the dosage and told the GP that the patient should be given that dose of X. The GP said "My patient cannot take X, because it harms her." The secretary asked the consultant what to do. "Allocate the next drug on the list that is not X."

A forestry experiment in a rectangle

7 varieties of guayule tree in a 5×7 rectangle, using a randomized complete-block design with the rows as blocks.

B	D	G	A	F	C	E
A	G	C	D	F	B	E
G	E	D	F	B	C	A
B	A	C	F	G	E	D
G	B	F	C	D	A	E

A forestry experiment in a rectangle

7 varieties of guayule tree in a 5×7 rectangle, using a randomized complete-block design with the rows as blocks.

B	D	G	A	F	C	E
A	G	C	D	F	B	E
G	E	D	F	B	C	A
B	A	C	F	G	E	D
G	B	F	C	D	A	E

A forestry experiment in a rectangle

7 varieties of guayule tree in a 5×7 rectangle, using a randomized complete-block design with the rows as blocks.

B	D	G	A	F	C	E
A	G	C	D	F	B	E
G	E	D	F	B	C	A
B	A	C	F	G	E	D
G	B	F	C	D	A	E

"Throw it away and re-randomize."

A forestry experiment in a rectangle

7 varieties of guayule tree in a 5×7 rectangle, using a randomized complete-block design with the rows as blocks.

B	D	G	A	F	C	E
A	G	C	D	F	B	E
G	E	D	F	B	C	A
B	A	C	F	G	E	D
G	B	F	C	D	A	E

"Throw it away and re-randomize."
For the 5×7 rectangle, the proportion of plans with no repeat in any column is only 0.000006.

Be honest with the statistician

"I didn't want to bother you with those details."

Constraints on the conduct of the experiment
should be incorporated into the design
(and therefore into the analysis),
not fudged in the randomization.

Replication

- If there is too much replication then the experiment may waste time and money. If animals are to be sacrificed, it is unethical to use too many.

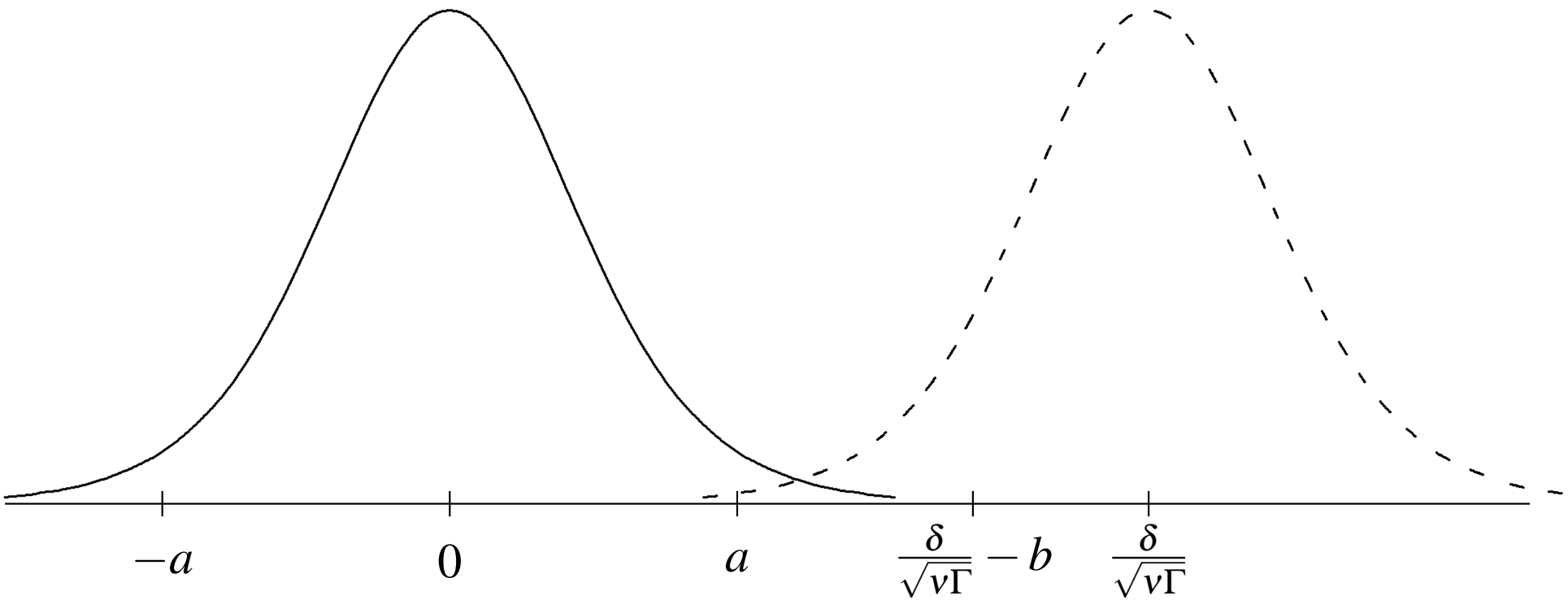
Replication

- If there is too much replication then the experiment may waste time and money. If animals are to be sacrificed, it is unethical to use too many.
- If there is too little replication then any genuine differences between treatments may be masked by the differences among the experimental units. An experiment which is too small to give any conclusions is also a waste of resources. It is also an unethical use of animals or people.

Replication

- If there is too much replication then the experiment may waste time and money. If animals are to be sacrificed, it is unethical to use too many.
- If there is too little replication then any genuine differences between treatments may be masked by the differences among the experimental units. An experiment which is too small to give any conclusions is also a waste of resources. It is also an unethical use of animals or people.
- Watch out for false replication.

Replication for power (two treatments)



Solid curve defines the interval $[-a, a]$ used for the hypothesis test (where a depends on the significance level);
dashed curve gives the probability density function of the test statistic = difference/s.e.d. $= \Delta/\sqrt{\nu\Gamma}$ if the real difference is δ;
$\Delta =$ estimate of δ; $\Gamma =$ estimate of variance per response;
$\nu =$ sum of reciprocals of replications; b defines the power.

False replication

Three pesticides were compared for their side-effects on ladybirds.

A field was divided into three areas and one pesticide applied to each area. Ladybirds were counted on three samples from each area.

$$\begin{aligned} \text{Treatments} &= ? \\ \text{Experimental units} &= ? \\ \text{Observational units} &= ? \\ \text{Replication} &= ? \end{aligned}$$

False replication

Three pesticides were compared for their side-effects on ladybirds.

A field was divided into three areas and one pesticide applied to each area. Ladybirds were counted on three samples from each area.

$$\begin{aligned} \text{Treatments} &= 3 \text{ pesticides} \\ \text{Experimental units} &= ? \\ \text{Observational units} &= ? \\ \text{Replication} &= ? \end{aligned}$$

False replication

Three pesticides were compared for their side-effects on ladybirds.

A field was divided into three areas and one pesticide applied to each area. Ladybirds were counted on three samples from each area.

$$\begin{aligned} \text{Treatments} &= 3 \text{ pesticides} \\ \text{Experimental units} &= 3 \text{ areas} \\ \text{Observational units} &= ? \\ \text{Replication} &= ? \end{aligned}$$

False replication

Three pesticides were compared for their side-effects on ladybirds.

A field was divided into three areas and one pesticide applied to each area. Ladybirds were counted on three samples from each area.

$$\begin{aligned} \text{Treatments} &= 3 \text{ pesticides} \\ \text{Experimental units} &= 3 \text{ areas} \\ \text{Observational units} &= 9 \text{ samples} \\ \text{Replication} &= ? \end{aligned}$$

False replication

Three pesticides were compared for their side-effects on ladybirds.

A field was divided into three areas and one pesticide applied to each area. Ladybirds were counted on three samples from each area.

$$\begin{aligned} \text{Treatments} &= 3 \text{ pesticides} \\ \text{Experimental units} &= 3 \text{ areas} \\ \text{Observational units} &= 9 \text{ samples} \\ \text{Replication} &= 1 \end{aligned}$$

Calf-feeding experiment

Calves were housed in pens, with ten calves per pen. Each pen was allocated to a certain type of feed. Batches of this type of feed were put into the pen; calves were free to eat as much of this as they liked. Calves were weighed individually.

Feed D Pen 1 10 calves	Feed C Pen 2 10 calves	Feed D Pen 3 10 calves	Feed B Pen 4 10 calves
Feed B Pen 5 10 calves	Feed A Pen 6 10 calves	Feed A Pen 7 10 calves	Feed C Pen 8 10 calves

Calf-feeding experiment

Calves were housed in pens, with ten calves per pen. Each pen was allocated to a certain type of feed. Batches of this type of feed were put into the pen; calves were free to eat as much of this as they liked. Calves were weighed individually.

Feed D Pen 1 10 calves	Feed C Pen 2 10 calves	Feed D Pen 3 10 calves	Feed B Pen 4 10 calves
Feed B Pen 5 10 calves	Feed A Pen 6 10 calves	Feed A Pen 7 10 calves	Feed C Pen 8 10 calves

treatment $=$ type of feed

Calf-feeding experiment

Calves were housed in pens, with ten calves per pen. Each pen was allocated to a certain type of feed. Batches of this type of feed were put into the pen; calves were free to eat as much of this as they liked. Calves were weighed individually.

Feed D Pen 1 10 calves	Feed C Pen 2 10 calves	Feed D Pen 3 10 calves	Feed B Pen 4 10 calves
Feed B Pen 5 10 calves	Feed A Pen 6 10 calves	Feed A Pen 7 10 calves	Feed C Pen 8 10 calves

treatment $=$ type of feed experimental unit $=$ pen

Calf-feeding experiment

Calves were housed in pens, with ten calves per pen. Each pen was allocated to a certain type of feed. Batches of this type of feed were put into the pen; calves were free to eat as much of this as they liked. Calves were weighed individually.

Feed D Pen 1 10 calves	Feed C Pen 2 10 calves	Feed D Pen 3 10 calves	Feed B Pen 4 10 calves
Feed B Pen 5 10 calves	Feed A Pen 6 10 calves	Feed A Pen 7 10 calves	Feed C Pen 8 10 calves

treatment = type of feed experimental unit = pen

observational unit = calf

Calf-feeding experiment: analysis of variance

Feed D Pen 1 10 calves	Feed C Pen 2 10 calves	Feed D Pen 3 10 calves	Feed B Pen 4 10 calves
Feed B Pen 5 10 calves	Feed A Pen 6 10 calves	Feed A Pen 7 10 calves	Feed C Pen 8 10 calves

Calf-feeding experiment: analysis of variance

Feed D Pen 1 10 calves	Feed C Pen 2 10 calves	Feed D Pen 3 10 calves	Feed B Pen 4 10 calves
Feed B Pen 5 10 calves	Feed A Pen 6 10 calves	Feed A Pen 7 10 calves	Feed C Pen 8 10 calves

Stratum	Source	Degrees of freedom
mean	mean	1
pens	feed	3
	residual	4
	total	7
calves	calves	72
Total		80

Calf-feeding experiment: analysis of variance

Feed D Pen 1 10 calves	Feed C Pen 2 10 calves	Feed D Pen 3 10 calves	Feed B Pen 4 10 calves
Feed B Pen 5 10 calves	Feed A Pen 6 10 calves	Feed A Pen 7 10 calves	Feed C Pen 8 10 calves

Stratum	Source	Degrees of freedom
mean	mean	1
pens	feed	3
	residual	4 no matter how many calves per pen
	total	7
calves	calves	72
Total		80

Calf-feeding experiment: changing the numbers

If feeds can be allocated only to whole pens, then

Calf-feeding experiment: changing the numbers

If feeds can be allocated only to whole pens, then

- increasing the number of calves per pen
  decreases the variance but
  does not increase the number of degrees of freedom for residual

Calf-feeding experiment: changing the numbers

If feeds can be allocated only to whole pens, then

- increasing the number of calves per pen
  decreases the variance but
  does not increase the number of degrees of freedom for residual
- increasing the number of pens
  decreases the variance and also
  increases the number of degrees of freedom for residual.

Educating general practitioners

A trial was conducted to test the effectiveness of new guidelines for the treatment of diabetes. Some GPs were randomized to the 'intervention' treatment and asked to attend some educational sessions where the new guidelines were explained; the other GPs in the experiment were not invited to such sessions. However, the observational units in the experiment were the diabetic patients of the two sets of GPs.

Educating general practitioners

A trial was conducted to test the effectiveness of new guidelines for the treatment of diabetes. Some GPs were randomized to the 'intervention' treatment and asked to attend some educational sessions where the new guidelines were explained; the other GPs in the experiment were not invited to such sessions. However, the observational units in the experiment were the diabetic patients of the two sets of GPs.

treatment	experimental unit	observational unit
feed	pen	calf

Educating general practitioners

A trial was conducted to test the effectiveness of new guidelines for the treatment of diabetes. Some GPs were randomized to the 'intervention' treatment and asked to attend some educational sessions where the new guidelines were explained; the other GPs in the experiment were not invited to such sessions. However, the observational units in the experiment were the diabetic patients of the two sets of GPs.

treatment	experimental unit	observational unit
feed	pen	calf
guidelines/not	GP	patient

Educating general practitioners

A trial was conducted to test the effectiveness of new guidelines for the treatment of diabetes. Some GPs were randomized to the 'intervention' treatment and asked to attend some educational sessions where the new guidelines were explained; the other GPs in the experiment were not invited to such sessions. However, the observational units in the experiment were the diabetic patients of the two sets of GPs.

treatment	experimental unit	observational unit
feed	pen	calf
guidelines/not	GP	patient
??	cage	rat

Regenerating bone

A biomaterials scientist is interested in the properties of ceramic scaffolds. These materials have the potential to regenerate bones in humans who have lost bone matter because of disease or trauma. The regulatory authorities demand that the materials be tested for efficacy and safety before being tried in humans, so he experiments on dogs, using two ceramic scaffolds and a 'do nothing' control. A portion of bone is damaged; the treatment is applied and left for several weeks; then the dog is killed so that the bone can be extracted, examined and weighed.

Regenerating bone

A biomaterials scientist is interested in the properties of ceramic scaffolds. These materials have the potential to regenerate bones in humans who have lost bone matter because of disease or trauma. The regulatory authorities demand that the materials be tested for efficacy and safety before being tried in humans, so he experiments on dogs, using two ceramic scaffolds and a 'do nothing' control. A portion of bone is damaged; the treatment is applied and left for several weeks; then the dog is killed so that the bone can be extracted, examined and weighed.

By using three bones per dog rather than one, he is able to obtain useful information by using only one quarter of the number of dogs envisaged when it was planned to treat only one bone per dog.

Regenerating bone

A biomaterials scientist is interested in the properties of ceramic scaffolds. These materials have the potential to regenerate bones in humans who have lost bone matter because of disease or trauma. The regulatory authorities demand that the materials be tested for efficacy and safety before being tried in humans, so he experiments on dogs, using two ceramic scaffolds and a 'do nothing' control. A portion of bone is damaged; the treatment is applied and left for several weeks; then the dog is killed so that the bone can be extracted, examined and weighed.

By using three bones per dog rather than one, he is able to obtain useful information by using only one quarter of the number of dogs envisaged when it was planned to treat only one bone per dog.

Within-dog variability should be less than between-dog variability: exploit this.

Design of the TeGenero trial

First-in-Man trial of a monoclonal antibody on healthy volunteers, March 2006: 4 cohorts of 8 volunteers each.

Cohort	TGN1412		Placebo
	Dose mg/kg body-weight	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

What happened to Cohort 1 on 13 March 2006

Healthy Volunteer	Randomised to	Time of intravenous administration	Time of transfer to critical care
A	TGN1412 8.4mg	0800	2400
B	Placebo	0810	
C	TGN1412 6.8mg	0820	2350
D	TGN1412 8.8mg	0830	0030
E	TGN1412 8.2mg	0840	2040
F	TGN1412 7.2mg	0850	0050
G	TGN1412 8.2mg	0900	0100
H	Placebo	0910	

The Royal Statistical Society's Working Party on Statistical Issues in First-in-Man Studies: Membership

Dipti Amin, Senior Vice-President, Quintiles
R. A. Bailey, Professor of Statistics, QMUL
Sheila Bird, Principal Scientist/Statistician, MRC Biostatistics Unit
Barbara Bogacka, Reader in Probability and Statistics, QMUL
Peter Colman, Senior Consultant Statistician, Pfizer
Andrew Garrett, Vice-President Statistics, Quintiles
Andrew Grieve, Professor of Medical Statistics, KCL
Peter Lachmann, FRS, Emeritus Professor of Immunology, Cambridge
Stephen Senn, Professor of Statistics, Glasgow

Planned analysis of the TeGenero trial

Cohort	TGN1412		Placebo
	Dose	Number	Number
1	1	6	2
2	2	6	2
3	3	6	2
4	4	6	2

If all responses are uncorrelated with variance σ^2 then
Variance (dose $i-$ placebo) in cohort i is $\left(\frac{1}{6}+\frac{1}{2}\right)\sigma^2 = \frac{2}{3}\sigma^2$

Planned analysis of the TeGenero trial

Cohort	TGN1412		Placebo
	Dose	Number	Number
1	1	6	2
2	2	6	2
3	3	6	2
4	4	6	2

If all responses are uncorrelated with variance σ^2 then
Variance (dose $i-$ placebo) in cohort i is $\left(\frac{1}{6}+\frac{1}{2}\right)\sigma^2 = \frac{2}{3}\sigma^2$

From the protocol: "data of subjects having received placebo will be pooled in one group for analyses."

Planned analysis of the TeGenero trial

Cohort	TGN1412		Placebo
	Dose	Number	Number
1	1	6	2
2	2	6	2
3	3	6	2
4	4	6	2

If all responses are uncorrelated with variance σ^2 then
Variance (dose $i-$ placebo) in cohort i is $\left(\frac{1}{6}+\frac{1}{2}\right)\sigma^2 = \frac{2}{3}\sigma^2$

From the protocol: "data of subjects having received placebo will be pooled in one group for analyses."

Variance (dose $i-$ placebo) is $\left(\frac{1}{6}+\frac{1}{8}\right)\sigma^2 = \frac{7}{24}\sigma^2$

Planned analysis of the TeGenero trial

Cohort	TGN1412		Placebo
	Dose	Number	Number
1	1	6	2
2	2	6	2
3	3	6	2
4	4	6	2

If all responses are uncorrelated with variance σ^2 then
Variance (dose $i-$ placebo) in cohort i is $\left(\frac{1}{6}+\frac{1}{2}\right)\sigma^2 = \frac{2}{3}\sigma^2$

From the protocol: "data of subjects having received placebo will be pooled in one group for analyses."

Variance (dose $i-$ placebo) is $\left(\frac{1}{6}+\frac{1}{8}\right)\sigma^2 = \frac{7}{24}\sigma^2$ if there are no cohort effects.

Planned analysis of the TeGenero trial

Cohort	TGN1412		Placebo
	Dose	Number	Number
1	1	6	2
2	2	6	2
3	3	6	2
4	4	6	2

If all responses are uncorrelated with variance σ^2 then
Variance (dose $i-$ placebo) in cohort i is $\left(\frac{1}{6}+\frac{1}{2}\right)\sigma^2 = \frac{2}{3}\sigma^2$

From the protocol: "data of subjects having received placebo will be pooled in one group for analyses."

Variance (dose $i-$ placebo) is $\left(\frac{1}{6}+\frac{1}{8}\right)\sigma^2 = \frac{7}{24}\sigma^2$ if there are no cohort effects.

Variance (dose $i-$ dose j) is $\left(\frac{1}{6}+\frac{1}{6}\right)\sigma^2 = \frac{1}{3}\sigma^2$ if there are no cohort effects.

Are there cohort effects?

- Different types of people can volunteer at different times.

Are there cohort effects?

- Different types of people can volunteer at different times.
- There may be changes in the ambient conditions,
  eg temperature, pollutants, pollens.

Are there cohort effects?

- Different types of people can volunteer at different times.
- There may be changes in the ambient conditions,
  eg temperature, pollutants, pollens.
- The staff running the trial, or analysing the samples, may change.

Are there cohort effects?

- Different types of people can volunteer at different times.
- There may be changes in the ambient conditions,
  eg temperature, pollutants, pollens.
- The staff running the trial, or analysing the samples, may change.
- Protocols for using subsidiary equipment may change.

Are there cohort effects?

- Different types of people can volunteer at different times.
- There may be changes in the ambient conditions,
  eg temperature, pollutants, pollens.
- The staff running the trial, or analysing the samples, may change.
- Protocols for using subsidiary equipment may change.
- Halo effect among volunteers:
  if one reports nausea then they all may do so.

Are there cohort effects?

- Different types of people can volunteer at different times.
- There may be changes in the ambient conditions,
  eg temperature, pollutants, pollens.
- The staff running the trial, or analysing the samples, may change.
- Protocols for using subsidiary equipment may change.
- Halo effect among volunteers:
  if one reports nausea then they all may do so.
- Halo effect among staff:
  if they see symptoms in one volunteer, they expect them in others.

Are there cohort effects?

- Different types of people can volunteer at different times.
- There may be changes in the ambient conditions,
  eg temperature, pollutants, pollens.
- The staff running the trial, or analysing the samples, may change.
- Protocols for using subsidiary equipment may change.
- Halo effect among volunteers:
  if one reports nausea then they all may do so.
- Halo effect among staff:
  if they see symptoms in one volunteer, they expect them in others.

Are there cohort effects?

- Different types of people can volunteer at different times.
- There may be changes in the ambient conditions,
  eg temperature, pollutants, pollens.
- The staff running the trial, or analysing the samples, may change.
- Protocols for using subsidiary equipment may change.
- Halo effect among volunteers:
  if one reports nausea then they all may do so.
- Halo effect among staff:
  if they see symptoms in one volunteer, they expect them in others.

There have been many trials, in many topics, where, with hindsight, cohort effects swamp treatment effects.

Are there cohort effects?

- Different types of people can volunteer at different times.
- There may be changes in the ambient conditions,
  eg temperature, pollutants, pollens.
- The staff running the trial, or analysing the samples, may change.
- Protocols for using subsidiary equipment may change.
- Halo effect among volunteers:
  if one reports nausea then they all may do so.
- Halo effect among staff:
  if they see symptoms in one volunteer, they expect them in others.

There have been many trials, in many topics, where, with hindsight, cohort effects swamp treatment effects.
The Experimental Medicines Group of the Association of the British Pharmaceutical Industry (ABPI) says that trials should always be designed on the assumption that there will be cohort effects.

Analysis of the TeGenero trial with cohort effects

Cohort	TGN1412		Placebo
	Dose	Number	Number
1	1	6	2
2	2	6	2
3	3	6	2
4	4	6	2

$$\text{Variance (dose } i - \text{ placebo) in cohort } i = \left(\frac{1}{6} + \frac{1}{2}\right)\sigma^2 = \frac{2}{3}\sigma^2.$$

Analysis of the TeGenero trial with cohort effects

Cohort	TGN1412		Placebo
	Dose	Number	Number
1	1	6	2
2	2	6	2
3	3	6	2
4	4	6	2

$$\text{Variance (dose } i - \text{ placebo) in cohort } i = \left(\frac{1}{6} + \frac{1}{2}\right)\sigma^2 = \frac{2}{3}\sigma^2.$$

$$\begin{aligned}&\text{Estimator of (dose } i - \text{ dose } j) = \\ &\quad [\text{estimator of (dose } i - \text{ placebo) in cohort } i] - \\ &\quad [\text{estimator of (dose } j - \text{ placebo) in cohort } j]\end{aligned}$$

Analysis of the TeGenero trial with cohort effects

Cohort	TGN1412		Placebo
	Dose	Number	Number
1	1	6	2
2	2	6	2
3	3	6	2
4	4	6	2

$$\text{Variance (dose } i - \text{ placebo) in cohort } i = \left(\frac{1}{6} + \frac{1}{2}\right)\sigma^2 = \frac{2}{3}\sigma^2.$$

Estimator of (dose $i -$ dose j) =

[estimator of (dose $i -$ placebo) in cohort i] $-$

[estimator of (dose $j -$ placebo) in cohort j]

$$\text{So variance (dose } i - \text{ dose } j) = \left(\frac{2}{3} + \frac{2}{3}\right)\sigma^2 = \frac{4}{3}\sigma^2.$$

Two designs for 4 doses using 40 subjects

Numbers of subjects

Old

Dose	0	1	2	3	4
Cohort 1	2	8	0	0	0
Cohort 2	2	0	8	0	0
Cohort 3	2	0	0	8	0
Cohort 4	2	0	0	0	8

Actual pairwise variances/σ^2

	1	2	3	4
0	0.625	0.625	0.625	0.625
1		1.250	1.250	1.250
2			1.250	1.250
3				1.250

New

Dose	0	1	2	3	4
Cohort 1	4	4	0	0	0
Cohort 2	2	2	4	0	0
Cohort 3	1	1	2	4	0
Cohort 4	1	1	1	1	4
Cohort 5	1	1	1	2	3

	1	2	3	4
0	0.222	0.285	0.348	0.370
1		0.285	0.348	0.370
2			0.330	0.378
3				0.375

Two designs for 4 doses using 40 subjects

Numbers of subjects

Old

Dose	0	1	2	3	4
Cohort 1	2	8	0	0	0
Cohort 2	2	0	8	0	0
Cohort 3	2	0	0	8	0
Cohort 4	2	0	0	0	8

Actual pairwise variances/σ^2

	1	2	3	4
0	0.625	0.625	0.625	0.625
1		1.250	1.250	1.250
2			1.250	1.250
3				1.250

average 1.00

New

Dose	0	1	2	3	4
Cohort 1	4	4	0	0	0
Cohort 2	2	2	4	0	0
Cohort 3	1	1	2	4	0
Cohort 4	1	1	1	1	4
Cohort 5	1	1	1	2	3

	1	2	3	4
0	0.222	0.285	0.348	0.370
1		0.285	0.348	0.370
2			0.330	0.378
3				0.375

average 0.33

References

- R. A. Bailey: *Design of Comparative Experiments*, Cambridge University Press, Cambridge, 2008.
- S. Senn, D. Amin, R. A. Bailey, S. M. Bird, B. Bogacka, P. Colman, A. Garrett, A. Grieve and P. Lachmann: Statistical issues in first-in-man studies. *Journal of the Royal Statistical Society, Series A* **170** (2007), 517–579.
- R. A. Bailey: Designs for dose-escalation trials with quantitative responses. *Statistics in Medicine* (2009), in press.