Randomization 50 years after Fisher

R. A. Bailey

r.a.bailey@qmul.ac.uk

Royal Statistical Society Conference,
September 2012

Abstract

One of the most important ideas that R. A. Fisher introduced into experimentation during his time at Rothamsted Experimental Station was randomisation.

Most people agree with that.

However, it turns out that they disagree about what is meant by randomisation: what it is, how you should do it, what its purpose is, whether or not it is desirable, and so on.

I shall try to cover some of the different points of view.

A lady tasting tea

" A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup." (Fisher, *Design of Experiments*, 1935)

A lady tasting tea

" A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup." (Fisher, *Design of Experiments*, 1935)

An experiment to test her assertion has

- 8 experimental units—8 cups in order
- 2 treatments—milk first or milk second.

A lady tasting tea

" A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup." (Fisher, *Design of Experiments*, 1935)

An experiment to test her assertion has

- 8 experimental units—8 cups in order
- 2 treatments—milk first or milk second.

How should the 2 treatments (call them A and B for short) be allocated to the 8 cups?

Two possible methods of randomization

- Start with the systematic plan *AAAABBBB*
  and then permute it by a permutation of 8 objects
  chosen at random from the set of all 8! such permutations
  —equivalently, choose at random from the 70 sequences of
  4 *A*s and 4 *B*s.

Two possible methods of randomization

- Start with the systematic plan $AAAABBBB$
  and then permute it by a permutation of 8 objects
  chosen at random from the set of all 8! such permutations
  —equivalently, choose at random from the 70 sequences of
  4 As and 4 Bs.
- For each cup independently, toss a fair coin
  and allocate A if it falls heads, B if it falls tails
  —equivalently, choose at random from the 256 binary
  sequences of length 8.

Two possible methods of randomization

- Start with the systematic plan $AAAABBBB$
  and then permute it by a permutation of 8 objects
  chosen at random from the set of all 8! such permutations
  —equivalently, choose at random from the 70 sequences of
  4 As and 4 Bs.
- For each cup independently, toss a fair coin
  and allocate A if it falls heads, B if it falls tails
  —equivalently, choose at random from the 256 binary
  sequences of length 8.

Two possible methods of randomization

- Start with the systematic plan *AAAABBBB*
  and then permute it by a permutation of 8 objects
  chosen at random from the set of all 8! such permutations
  —equivalently, choose at random from the 70 sequences of
  4 *A*s and 4 *B*s.
- For each cup independently, toss a fair coin
  and allocate *A* if it falls heads, *B* if it falls tails
  —equivalently, choose at random from the 256 binary
  sequences of length 8.

If she has no discrimination, what is the chance that she will identify each cup correctly?

Two possible methods of randomization

- Start with the systematic plan *AAAABBBB*
  and then permute it by a permutation of 8 objects
  chosen at random from the set of all 8! such permutations
  —equivalently, choose at random from the 70 sequences of
  4 *A*s and 4 *B*s.
- For each cup independently, toss a fair coin
  and allocate *A* if it falls heads, *B* if it falls tails
  —equivalently, choose at random from the 256 binary
  sequences of length 8.

If she has no discrimination, what is the chance that she will identify each cup correctly?

Does this depend on her knowing the method of randomization?

The lady tasting tea is a red herring

Most experiments are not like this.

Usually, we want to

- estimate the difference between A and B
- estimate the standard error of this difference
- perform a hypothesis test to see whether A and B differ.

The lady tasting tea is a red herring

Most experiments are not like this.

Usually, we want to

- estimate the difference between A and B
- estimate the standard error of this difference
- perform a hypothesis test to see whether A and B differ.

. . . and there may be more than 2 treatments.

Three examples: 5 treatments in 50 experimental units

1. I have 5 varieties of wheat to compare, using 50 plots in a single field. I will plant them all at the same time, and I can find out a lot of information about the plots beforehand. It is possible (but not very likely) that I may lose some plots to flooding or pests or mis-management.

Three examples: 5 treatments in 50 experimental units

1. I have 5 varieties of wheat to compare, using 50 plots in a single field. I will plant them all at the same time, and I can find out a lot of information about the plots beforehand. It is possible (but not very likely) that I may lose some plots to flooding or pests or mis-management.
2. I have 5 biological substances that I want to compare in the lab. I have 10 samples of each. The 50 procedures must be done one at a time, and it will take me a week to complete them all.

Three examples: 5 treatments in 50 experimental units

1. I have 5 varieties of wheat to compare, using 50 plots in a single field. I will plant them all at the same time, and I can find out a lot of information about the plots beforehand. It is possible (but not very likely) that I may lose some plots to flooding or pests or mis-management.
2. I have 5 biological substances that I want to compare in the lab. I have 10 samples of each. The 50 procedures must be done one at a time, and it will take me a week to complete them all.
3. I have 5 new therapies to compare, using 50 patients. Patients will be recruited sequentially, each one allocated to a therapy at recruitment. I do not know anything about these patients in advance, apart from the recruitment criteria. The trial may have to stop before 50 patients are recruited.

Some methods of randomization

- Randomly select the treatment for the next experimental unit, one unit at a time:

Some methods of randomization

- Randomly select the treatment for the next experimental unit, one unit at a time:
  - if these are all done independently, it is likely that replications will be unequal;

Some methods of randomization

- Randomly select the treatment for the next experimental unit, one unit at a time:
  - if these are all done independently, it is likely that replications will be unequal;
  - in sequential experiments, there are systems for biasing the next allocation depending on the allocations so far.

Some methods of randomization

- Randomly select the treatment for the next experimental unit, one unit at a time:
  - if these are all done independently, it is likely that replications will be unequal;
  - in sequential experiments, there are systems for biasing the next allocation depending on the allocations so far.
- Create a systematic plan
  and apply a permutation chosen at random from a suitable group of permutations of the experimental units.

Some methods of randomization

- Randomly select the treatment for the next experimental unit, one unit at a time:
  - if these are all done independently, it is likely that replications will be unequal;
  - in sequential experiments, there are systems for biasing the next allocation depending on the allocations so far.
- Create a systematic plan

  and apply a permutation chosen at random from a suitable group of permutations of the experimental units.
- Randomize the treatment labels:

Some methods of randomization

- Randomly select the treatment for the next experimental unit, one unit at a time:
  - if these are all done independently, it is likely that replications will be unequal;
  - in sequential experiments, there are systems for biasing the next allocation depending on the allocations so far.
- Create a systematic plan

  and apply a permutation chosen at random from a suitable group of permutations of the experimental units.
- Randomize the treatment labels:
  - this is not usually effective unless combined with the previous method;

Some methods of randomization

- Randomly select the treatment for the next experimental unit, one unit at a time:
  - if these are all done independently, it is likely that replications will be unequal;
  - in sequential experiments, there are systems for biasing the next allocation depending on the allocations so far.
- Create a systematic plan

  and apply a permutation chosen at random from a suitable group of permutations of the experimental units.
- Randomize the treatment labels:
  - this is not usually effective unless combined with the previous method;
  - it is not appropriate if some contrasts are deliberately intended to have higher precision that others.

Some methods of randomization

- Randomly select the treatment for the next experimental unit, one unit at a time:
  - if these are all done independently, it is likely that replications will be unequal;
  - in sequential experiments, there are systems for biasing the next allocation depending on the allocations so far.
- Create a systematic plan

  and apply a permutation chosen at random from a suitable group of permutations of the experimental units.
- Randomize the treatment labels:
  - this is not usually effective unless combined with the previous method;
  - it is not appropriate if some contrasts are deliberately intended to have higher precision that others.
- Choose at random from set of plans:

Some methods of randomization

- Randomly select the treatment for the next experimental unit, one unit at a time:
  - if these are all done independently, it is likely that replications will be unequal;
  - in sequential experiments, there are systems for biasing the next allocation depending on the allocations so far.
- Create a systematic plan
  and apply a permutation chosen at random from a suitable group of permutations of the experimental units.
- Randomize the treatment labels:
  - this is not usually effective unless combined with the previous method;
  - it is not appropriate if some contrasts are deliberately intended to have higher precision that others.
- Choose at random from set of plans:
  - all of the foregoing can be considered as special cases of this;

Some methods of randomization

- Randomly select the treatment for the next experimental unit, one unit at a time:
  - if these are all done independently, it is likely that replications will be unequal;
  - in sequential experiments, there are systems for biasing the next allocation depending on the allocations so far.
- Create a systematic plan

  and apply a permutation chosen at random from a suitable group of permutations of the experimental units.
- Randomize the treatment labels:
  - this is not usually effective unless combined with the previous method;
  - it is not appropriate if some contrasts are deliberately intended to have higher precision that others.
- Choose at random from set of plans:
  - all of the foregoing can be considered as special cases of this;
  - there are some sets of plans that cannot be obtained by the previous methods.

Randomization to avoid bias

One purpose of randomization 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)

Randomization to avoid bias

One purpose of randomization 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)

Randomization to avoid bias

One purpose of randomization 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)

Randomization to avoid bias

One purpose of randomization 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 or other people involved.

Randomization to avoid bias

One purpose of randomization 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 or other people involved.

Randomization to avoid bias

One purpose of randomization 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 or other people involved.

This must be done in a publicly convincing way (Cox, 2009).

Lanarkshire milk experiment: an example of cheating

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: an example of selection bias

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

Doctor knows best: an example of selection bias

A consultant organized a trial of 3 treatments to cure a serious disease: the current standard drug X, which was a very strong antibiotic, and 2 new drugs. Several GPs agreed to participate. They were sent the trial protocol, and asked to phone the consultant's secretary when they had a suitable patient. 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. The secretary asked 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 dose 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: an example of selection bias

A consultant organized a trial of 3 treatments to cure a serious disease: the current standard drug X, which was a very strong antibiotic, and 2 new drugs. Several GPs agreed to participate. They were sent the trial protocol, and asked to phone the consultant's secretary when they had a suitable patient. 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. The secretary asked 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 dose 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."

Random choice from exactly two plans

Two biologists investigated the effect of 2 different environments for female flour beetles in the 30 minutes after mating: what difference did this make to fertilization?
They used 74 female flour beetles. They tossed a coin to choose the first treatment, and then alternated them.

Random choice from exactly two plans

Two biologists investigated the effect of 2 different environments for female flour beetles in the 30 minutes after mating: what difference did this make to fertilization?
They used 74 female flour beetles. They tossed a coin to choose the first treatment, and then alternated them.

An educational psychologist compared 2 different methods of presenting information. Her experimental units were 30 undergraduates, who volunteered sequentially by arriving at her office. She tossed a coin to choose the first treatment, and then alternated them.

Random choice from exactly two plans

Two biologists investigated the effect of 2 different environments for female flour beetles in the 30 minutes after mating: what difference did this make to fertilization?
They used 74 female flour beetles. They tossed a coin to choose the first treatment, and then alternated them.

An educational psychologist compared 2 different methods of presenting information. Her experimental units were 30 undergraduates, who volunteered sequentially by arriving at her office. She tossed a coin to choose the first treatment, and then alternated them.

In both cases, the experimenter(s) knew the next treatment. Did this subconsciously affect their choice of next beetle or student?

Random choice from exactly two plans

Two biologists investigated the effect of 2 different environments for female flour beetles in the 30 minutes after mating: what difference did this make to fertilization?
They used 74 female flour beetles. They tossed a coin to choose the first treatment, and then alternated them.

An educational psychologist compared 2 different methods of presenting information. Her experimental units were 30 undergraduates, who volunteered sequentially by arriving at her office. She tossed a coin to choose the first treatment, and then alternated them.

In both cases, the experimenter(s) knew the next treatment. Did this subconsciously affect their choice of next beetle or student?

In the educational experiment, the students would have been able to spot the simple pattern. Did they deliberately volunteer in an order to get their chosen method?

Randomization to elicit more of the truth

1. Statistician and scientist discuss the planned experiment

   ...

Randomization to elicit more of the truth

1. Statistician and scientist discuss the planned experiment ...
2. ... and agree on a design.

Randomization to elicit more of the truth

1. Statistician and scientist discuss the planned experiment
   ...
2. ... and agree on a design.
3. Statistician randomizes the design to produce a field plan.

Randomization to elicit more of the truth

1. Statistician and scientist discuss the planned experiment …
2. … and agree on a design.
3. Statistician randomizes the design to produce a field plan.
4. Scientist says "Oh, I can't possibly do it that way because …"

Better ways to remove bias: Diffusion of proteins

A post-doc added from 0 to 4 extra green fluorescent proteins to cells of *Escherichia coli*, adding 0 to 10 cells, 1 to 10 further cells, and so on. Then she measured the rate of diffusion of proteins in each of the 50 cells.

Better ways to remove bias: Diffusion of proteins

A post-doc added from 0 to 4 extra green fluorescent proteins to cells of *Escherichia coli*, adding 0 to 10 cells, 1 to 10 further cells, and so on. Then she measured the rate of diffusion of proteins in each of the 50 cells.

This is what she did.

Monday	Tuesday	Wednesday	Thursday	Friday
0000000000	1111111111	2222222222	3333333333	4444444444

Better ways to remove bias: Diffusion of proteins

A post-doc added from 0 to 4 extra green fluorescent proteins to cells of *Escherichia coli*, adding 0 to 10 cells, 1 to 10 further cells, and so on. Then she measured the rate of diffusion of proteins in each of the 50 cells.

This is what she did.

Monday	Tuesday	Wednesday	Thursday	Friday
0000000000	1111111111	2222222222	3333333333	4444444444

Are the perceived differences caused by differences in size?

Better ways to remove bias: Diffusion of proteins

A post-doc added from 0 to 4 extra green fluorescent proteins to cells of *Escherichia coli*, adding 0 to 10 cells, 1 to 10 further cells, and so on. Then she measured the rate of diffusion of proteins in each of the 50 cells.

This is what she did.

Monday	Tuesday	Wednesday	Thursday	Friday
0000000000	1111111111	2222222222	3333333333	4444444444

Are the perceived differences caused by differences in size?

Did she get better at preparing the samples as the week wore on?

Better ways to remove bias: Diffusion of proteins

A post-doc added from 0 to 4 extra green fluorescent proteins to cells of *Escherichia coli*, adding 0 to 10 cells, 1 to 10 further cells, and so on. Then she measured the rate of diffusion of proteins in each of the 50 cells.

This is what she did.

Monday	Tuesday	Wednesday	Thursday	Friday
0000000000	1111111111	2222222222	3333333333	4444444444

Are the perceived differences caused by differences in size?

Did she get better at preparing the samples as the week wore on?

Were there environmental changes in the lab that could have contributed to the differences?

Diffusion of proteins: continued

What she did.

Monday	Tuesday	Wednesday	Thursday	Friday
0000000000	1111111111	2222222222	3333333333	4444444444

Diffusion of proteins: continued

What she did.

Monday	Tuesday	Wednesday	Thursday	Friday
0000000000	1111111111	2222222222	3333333333	4444444444

Better to regard each day as a block.

Monday	Tuesday	Wednesday	Thursday	Friday
0011223344	0011223344	0011223344	0011223344	0011223344

Diffusion of proteins: continued

What she did.

Monday	Tuesday	Wednesday	Thursday	Friday
0000000000	1111111111	2222222222	3333333333	4444444444

Better to regard each day as a block.

Monday	Tuesday	Wednesday	Thursday	Friday
0011223344	0011223344	0011223344	0011223344	0011223344

There may still be systematic differences within each day,
so better still, randomize within each day.

Diffusion of proteins: continued

What she did.

Monday	Tuesday	Wednesday	Thursday	Friday
0000000000	1111111111	2222222222	3333333333	4444444444

Better to regard each day as a block.

Monday	Tuesday	Wednesday	Thursday	Friday
0011223344	0011223344	0011223344	0011223344	0011223344

There may still be systematic differences within each day,
so better still, randomize within each day.

Monday	Tuesday	Wednesday	Thursday	Friday
1040223134	2230110443	1421324030	4420013312	3204320411

Blocking

Partition the experimental units into blocks in such a way that units within a block are more alike than units in different blocks.

Blocking

Partition the experimental units into blocks in such a way that units within a block are more alike than units in different blocks.

Choose how to allocate treatments to units: in simple cases, all treatments appear equally often in each block. This should remove known sources of bias.

Blocking

Partition the experimental units into blocks in such a way that units within a block are more alike than units in different blocks.

Choose how to allocate treatments to units: in simple cases, all treatments appear equally often in each block. This should remove known sources of bias.

Randomize by

- randomizing the labels of the blocks (this has no effect if they all have the same treatments);
- within each block independently, randomizing the order of the units.

This should remove unknown sources of bias.

Blocking

Partition the experimental units into blocks in such a way that units within a block are more alike than units in different blocks.

Choose how to allocate treatments to units: in simple cases, all treatments appear equally often in each block. This should remove known sources of bias.

Randomize by

- randomizing the labels of the blocks (this has no effect if they all have the same treatments);
- within each block independently, randomizing the order of the units.

This should remove unknown sources of bias.

Allow for the blocks in the data analysis. If you do not do this, you over-estimate the error variance.

Additivity

Assume that the responses on the experimental units are measured on such a scale that, if unit ω is allocated treatment i then the response Y_ω satifies

$$Y_\omega = X_\omega + \tau_i,$$

where we cannot know X_ω but we want to know τ_i.
(Kempthorne, Why randomize?, 1977)

Additivity

Assume that the responses on the experimental units are measured on such a scale that, if unit ω is allocated treatment i then the response Y_ω satifies

$$Y_\omega = X_\omega + \tau_i,$$

where we cannot know X_ω but we want to know τ_i.
(Kempthorne, Why randomize?, 1977)

Too many parameters and variables!

Additivity

Assume that the responses on the experimental units are measured on such a scale that, if unit ω is allocated treatment i then the response Y_ω satifies

$$Y_\omega = X_\omega + \tau_i,$$

where we cannot know X_ω but we want to know τ_i.
(Kempthorne, Why randomize?, 1977)

Too many parameters and variables!

Does it matter whether we consider X_ω to be a constant or a random variable? If constant, do we need to add another (random) term for measurement error?
(Kempthorne, 1955; Bailey, 1991; Caliński and Kageyama, 2000).

One approach: randomization tests

If the treatment parameters τ_i are all equal, then any test statistic which is orthogonal to the all-1 vector is a function of the underlying unit values X_ω.

One approach: randomization tests

If the treatment parameters τ_i are all equal, then any test statistic which is orthogonal to the all-1 vector is a function of the underlying unit values X_ω.

Calculate this test statistic for all possible plans that can be produced by your method of randomization.
The more extreme is the original value within this distribution, the less likely it is that the treatment parameters τ_i are all equal.

One approach: randomization tests

If the treatment parameters τ_i are all equal, then any test statistic which is orthogonal to the all-1 vector is a function of the underlying unit values X_ω.

Calculate this test statistic for all possible plans that can be produced by your method of randomization.
The more extreme is the original value within this distribution, the less likely it is that the treatment parameters τ_i are all equal.

This approach views randomization as giving a basis for exact tests of significance without (many) model assumptions.

One approach: randomization tests

If the treatment parameters τ_i are all equal, then any test statistic which is orthogonal to the all-1 vector is a function of the underlying unit values X_ω.

Calculate this test statistic for all possible plans that can be produced by your method of randomization.
The more extreme is the original value within this distribution, the less likely it is that the treatment parameters τ_i are all equal.

This approach views randomization as giving a basis for exact tests of significance without (many) model assumptions.

(Edgington, *Randomization Tests*, 1987; Good, *Permutation Tests*, 1994)

Another approach: randomization justifies the model

Randomize by applying a randomly chosen permutation g from a suitable group G of permutations of the experimental units that preserves any structure such as blocking.

Another approach: randomization justifies the model

Randomize by applying a randomly chosen permutation g from a suitable group G of permutations of the experimental units that preserves any structure such as blocking.
Replace X_ω by the random variable Z_ω, where

$$\text{Prob}\left(Z_\omega = X_{g(\omega)} \text{ for all } \omega\right) = \frac{1}{|G|}$$

Another approach: randomization justifies the model

Randomize by applying a randomly chosen permutation g from a suitable group G of permutations of the experimental units that preserves any structure such as blocking.
Replace X_ω by the random variable Z_ω, where

$$\text{Prob}\left(Z_\omega = X_{g(\omega)} \text{ for all } \omega\right) = \frac{1}{|G|}$$

If G is transitive in the sense that, given any two units, there is at least one g in G taking one to the other,
then $\mathbb{E}(Z_\omega)$ does not depend on ω,
and so may be absorbed into all the treatment parameters,
and $\text{Var}(Z_\omega)$ does not depend on ω:

Another approach: randomization justifies the model

Randomize by applying a randomly chosen permutation g from a suitable group G of permutations of the experimental units that preserves any structure such as blocking.
Replace X_ω by the random variable Z_ω, where

$$\text{Prob}\left(Z_\omega = X_{g(\omega)} \text{ for all } \omega\right) = \frac{1}{|G|}$$

If G is transitive in the sense that, given any two units, there is at least one g in G taking one to the other,
then $\mathbb{E}(Z_\omega)$ does not depend on ω,
and so may be absorbed into all the treatment parameters,
and $\text{Var}(Z_\omega)$ does not depend on ω: put $\text{Var}(Z_\omega) = \sigma^2$ for all ω.

Another approach: randomization justifies the model

Randomize by applying a randomly chosen permutation g from a suitable group G of permutations of the experimental units that preserves any structure such as blocking.
Replace X_ω by the random variable Z_ω, where

$$\text{Prob}\left(Z_\omega = X_{g(\omega)} \text{ for all } \omega\right) = \frac{1}{|G|}$$

If G is transitive in the sense that, given any two units, there is at least one g in G taking one to the other,
then $\mathbb{E}(Z_\omega)$ does not depend on ω,
and so may be absorbed into all the treatment parameters,
and $\text{Var}(Z_\omega)$ does not depend on ω: put $\text{Var}(Z_\omega) = \sigma^2$ for all ω.

$$\text{Cov}(Z_\alpha, Z_\beta) = \text{Cov}(Z_{g(\alpha)}, Z_{g(\beta)})$$

for all units α and β, and all g in G.

Randomized blocks

If all blocks have the same size, and we randomize by using all permutations which preserve the partition into blocks, then our model becomes:
if unit ω is allocated treatment i, then

$$\mathbb{E}(Y_\omega) = \tau_i,$$

and

$$\mathrm{Cov}(Y_\alpha, Y_\beta) = \begin{cases} \sigma^2 & \text{if } \alpha = \beta \\ \rho_1\sigma^2 & \text{if } \alpha \neq \beta \text{ in the same block} \\ \rho_2\sigma^2 & \text{otherwise.} \end{cases}$$

Randomized blocks

If all blocks have the same size, and we randomize by using all permutations which preserve the partition into blocks, then our model becomes:
if unit ω is allocated treatment i, then

$$\mathbb{E}(Y_\omega) = \tau_i,$$

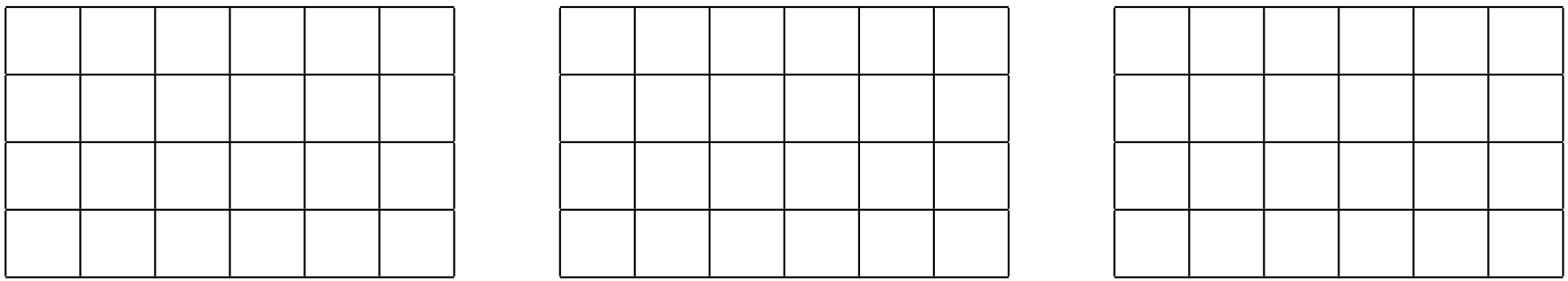
and

$$\mathrm{Cov}(Y_\alpha, Y_\beta) = \begin{cases} \sigma^2 & \text{if } \alpha = \beta \\ \rho_1\sigma^2 & \text{if } \alpha \neq \beta \text{ in the same block} \\ \rho_2\sigma^2 & \text{otherwise.} \end{cases}$$

The eigenspaces of the covariance matrix are the usual strata: grand mean, between blocks, and within-blocks.

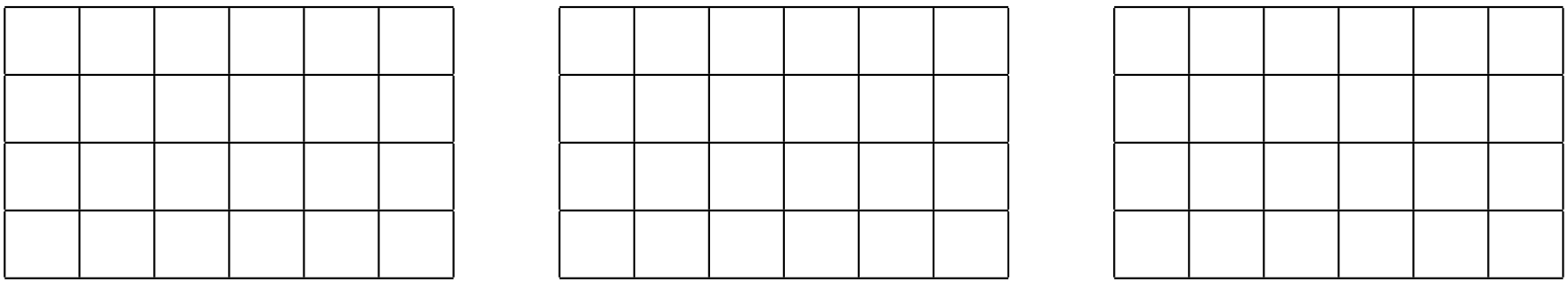
More complicated block structures

$$(3\text{ blocks})/((4\text{ rows}) \times (6\text{ columns}))$$



More complicated block structures

$$(3 \text{ blocks}) / ((4 \text{ rows}) \times (6 \text{ columns}))$$



Randomization:

- randomize the order of the blocks;
- within each block independently, randomize the order of the rows;
- within each block independently, randomize the order of the columns, independently of the order of the rows.

Simple orthogonal block structures

The crossing and nesting operators give rise to simple orthogonal block structures (Nelder, 1965), which have been generalized to poset block structures (Bailey, 2004).

Simple orthogonal block structures

The crossing and nesting operators give rise to simple orthogonal block structures (Nelder, 1965), which have been generalized to poset block structures (Bailey, 2004).

These are essentially the same as the complete balanced response structures of Kempthorne, Zyskind, Addelman, Throckmorton and White (1961), but this needs some proof, and their definition does not lend itself to the necessary algorithms.

Simple orthogonal block structures

The crossing and nesting operators give rise to simple orthogonal block structures (Nelder, 1965), which have been generalized to poset block structures (Bailey, 2004).

These are essentially the same as the complete balanced response structures of Kempthorne, Zyskind, Addelman, Throckmorton and White (1961), but this needs some proof, and their definition does not lend itself to the necessary algorithms.

Some non-trivial group theory shows that the randomization model for such structures gives a covariance matrix whose eigenspaces are precisely the strata usually used in the analysis of variance.

Another simple orthogonal block structure

0	160	240
160	80	80
80	0	160
240	240	0

↑ Cropper ↑ Melba ↑ Melle

160	80	0
0	160	80
240	0	240
80	240	160

↑ Melba ↑ Cropper ↑ Melle

experimental unit = plot

treatment = combination of cultivar and amount of fertilizer

Structure on the experimental units

(2 fields) / (3 strips) / (4 plots)

Structure on the experimental units

(2 fields) / (3 strips) / (4 plots)

Randomize fields; randomize strips within fields; randomize plots within strips.

Structure on the experimental units

(2 fields) / (3 strips) / (4 plots)

Randomize fields; randomize strips within fields; randomize plots within strips.

stratum	dim
overall mean	1
Fields	1
Strips[Fields]	4
Plots[Strips]	18

Structure on the experimental units

(2 fields) / (3 strips) / (4 plots)

Randomize fields; randomize strips within fields;
randomize plots within strips.

stratum	dim
overall mean	1
Fields	1
Strips[Fields]	4
Plots[Strips]	18

Some North Americans call this restricted randomization.

Statistical collaboration

1. Statistician and scientist discuss the planned experiment …
2. … and agree on a design.
3. Statistician randomizes the design to produce a field plan.
4. Scientist says "Oh, I can't possibly do it that way because …"

Statistical collaboration

1. Statistician and scientist discuss the planned experiment …
2. … and agree on a design.
3. Statistician randomizes the design to produce a field plan.
4. Scientist says "Oh, I can't possibly do it that way because …"

Do you

- Go back to step 3,
  rerandomize and hope that the next field plan will be OK?
  (but maybe you will reject a large proportion of plans)

Statistical collaboration

1. Statistician and scientist discuss the planned experiment …
2. … and agree on a design.
3. Statistician randomizes the design to produce a field plan.
4. Scientist says "Oh, I can't possibly do it that way because …"

Do you

- Go back to step 3,
  rerandomize and hope that the next field plan will be OK?
  (but maybe you will reject a large proportion of plans)
- Learn pertinent new information about constraints on the design, and so go back to step 1?

Statistical collaboration

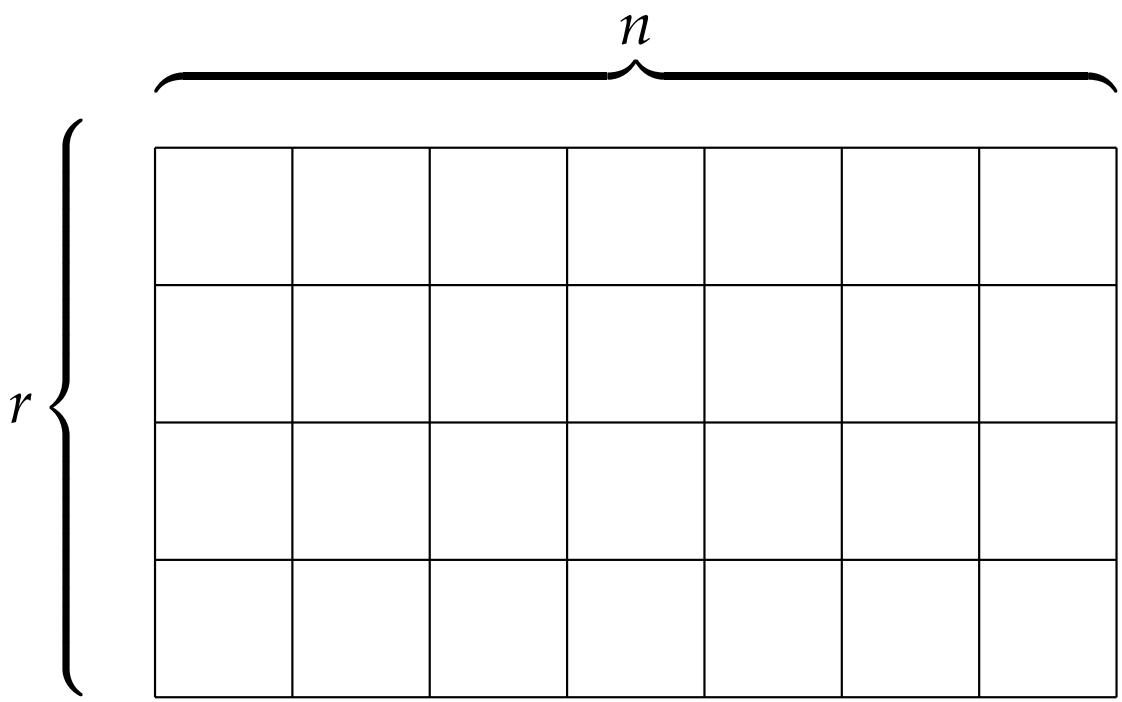
1. Statistician and scientist discuss the planned experiment …
2. … and agree on a design.
3. Statistician randomizes the design to produce a field plan.
4. Scientist says "Oh, I can't possibly do it that way because …"

Do you

- Go back to step 3,
  rerandomize and hope that the next field plan will be OK?
  (but maybe you will reject a large proportion of plans)
- Learn pertinent new information about constraints on the design, and so go back to step 1?
- Go back to step 2,
  and agree on a scheme of restricted randomization?

A problem in field trials

An agricultural experiment to compare n treatments.
The experimental area has r rows and n columns.



Use a randomized complete-block design with rows as blocks.
(In each row independently,
choose one of the $n!$ orders with equal probability.)

What should we do if the randomization produces a plan with one treatment always at one side of the rectangle?

Example

Federer (1955 book): guayule trees

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

Example

Federer (1955 book): guayule trees

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

Restricted randomization

Choose a special subset of permutations or of plans
which avoid certain bad patterns
while still giving unbiased estimators of treatment differences
and unbiased estimators of variance,
when averaged over all possible plans.

Restricted randomization

Choose a special subset of permutations or of plans
which avoid certain bad patterns
while still giving unbiased estimators of treatment differences
and unbiased estimators of variance,
when averaged over all possible plans.

Yates (1948); Grundy and Healy (1950); Bailey (1983).

Restricted randomization

Choose a special subset of permutations or of plans
which avoid certain bad patterns
while still giving unbiased estimators of treatment differences
and unbiased estimators of variance,
when averaged over all possible plans.

Yates (1948); Grundy and Healy (1950); Bailey (1983).

L. Moulton: talk on *Challenges in the design and analysis of a randomized, phased implementation (stepped-wedge) study in Brazil* at the Isaac Newton Institute in 2011, is using the criterion of validity proposed by this approach.

Super-valid restricted randomization

If you are willing to assume a little more about the underlying variables, it is possible to find schemes of restricted randomization for which the estimator of variance is unbiased when averaged over all comparisons in this one experiment.

Super-valid restricted randomization

If you are willing to assume a little more about the underlying variables, it is possible to find schemes of restricted randomization for which the estimator of variance is unbiased when averaged over all comparisons in this one experiment.

Bailey (2012)

Differences in approach: Latin squares

Here are some possible ways of randomizing a 5×5 Latin square. All give unbiased estimators of treatment differences and of variances.

Choose from all 5×5 Latin squares	161280
Start with a non-cyclic square; randomize rows, columns and letters	144000
Start with a cyclic square; randomize rows, columns and letters	17280
Start with a non-cyclic square; randomize rows and columns	2880
Start with a cyclic square; randomize rows and columns	2880
Choose one square at random from a complete set of 4 mutually orthogonal Latin squares; randomize letters	480

Differences in approach: Latin squares

Here are some possible ways of randomizing a 5×5 Latin square. All give unbiased estimators of treatment differences and of variances.

Choose from all 5×5 Latin squares	161280
Start with a non-cyclic square; randomize rows, columns and letters	144000
Start with a cyclic square; randomize rows, columns and letters	17280
Start with a non-cyclic square; randomize rows and columns	2880
Start with a cyclic square; randomize rows and columns	2880
Choose one square at random from a complete set of 4 mutually orthogonal Latin squares; randomize letters	480

Fisher insisted that only the first way is correct, but there may be an advantage in using a square that has an orthogonal mate.

Patients arriving sequentially

Efron's biased coin designs (1971)

Minimization: sequential balancing over many covariates (may have undesired side effects) (Pocock and Simon, 1975)

Biased coin with covariates (Atkinson, 1999)

Restricted randomization in random permuted blocks (Bailey and Nelson, 2003)

Other forms of restricted randomization (Plamadeala and Rosenberger, 2012)

Changing the randomization of later patients in the light of responses so far (Hu and Rosenberger, 2006; Coad, 2008)

Not in favour of randomization

W. S. Gosset ('Student') argued with Fisher in letters from 1915 to 1934.
Gosset claimed that 'balanced' designs, typically his ABBA designs, had smaller bias than completely randomized designs.

Not in favour of randomization

W. S. Gosset ('Student') argued with Fisher in letters from 1915 to 1934.
Gosset claimed that 'balanced' designs, typically his ABBA designs, had smaller bias than completely randomized designs.
The counter-argument was put eloquently by Fisher (1926) and Yates (1939): if there are patterns of variation in the field, and if treatments are allocated in such a way that these patterns have minimal effect on the estimate of treatment differences, then those patterns will inflate the estimates of standard errors of differences unless those patterns are allowed for in the data analysis.

Not in favour of randomization

W. S. Gosset ('Student') argued with Fisher in letters from 1915 to 1934.
Gosset claimed that 'balanced' designs, typically his ABBA designs, had smaller bias than completely randomized designs.
The counter-argument was put eloquently by Fisher (1926) and Yates (1939): if there are patterns of variation in the field, and if treatments are allocated in such a way that these patterns have minimal effect on the estimate of treatment differences, then those patterns will inflate the estimates of standard errors of differences unless those patterns are allowed for in the data analysis.

Ziliak (2011) supports Gosset's argument,
but he confuses 'plots' and 'blocks',
seems unaware of the possibilities for blocking in design and analysis,
and advocates false replication.

Randomization does not mean ignoring differences that you know about!

- Remove known sources of bias by using blocking or covariates. Design appropriately.
- Remove unknown sources of bias by randomizing appropriately.
- Allow for both of the above in the data analysis, so that estimates of treatment differences and their variances are unbiased.
- Do not overdo it: non-orthogonal designs give estimators with higher variance, and reduction in degrees of freedom reduces power.

References I

- A. C. Atkinson: Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statistics in Medicine* **18** (1999), 1741–52.
- R. A. Bailey: A unified approach to design of experiments. *JRSSA* **144** (1981), 214–223.
- R. A. Bailey: Restricted randomization. *Biometrika* **70** (1983), 183–198.
- R. A. Bailey: Strata for randomized experiments. *JRSSB* **53** (1991), 27–66.
- R. A. Bailey: *Association Schemes*. Cambridge University Press, 2004.
- R. A. Bailey: Experiments in rectangular areas: design and randomization. *JABES* **17** (2012), 176–191.
- R. A. Bailey and P. R. Nelson: Hadamard randomization: a valid restriction of random permuted blocks. *Biom. J.* **45** (2003), 554–560.

References II

- T. Caliński and S. Kageyama: *Block Designs: A Randomization Approach*. Springer, 2000.
- D. S. Coad: Response adaptive randomization. *Wiley Encycopedia of Clinical Trials*, 2008.
- D. R. Cox: Randomization in the design of experiments. *Int. Stat. Rev.* **77** (2009), 415–429.
- E. S. Edgington: *Randomization Tests*. Marcel Dekker, 1987.
- B. Efron. Forcing a sequential experiment to be balanced. *Biometrika* **58**, 403–417.
- W. T. Federer: *Experimental Design*. Macmillan, 1955.
- R. A. Fisher: The arrangement of field experiments. *J. Min. Agric. Gr. Br.* **33** (1926), 503–513.
- R. A. Fisher: *The Design of Experiments*. Oliver and Boyd, 1935.
- P. Good: *Permutation Tests*. Springer, 1994.
- P. M. Grundy and M. J. R. Healy: Restricted randomization and quasi-Latin squares. *JRSSB* **12** (1950), 286–291.

References III

- F. Hu and W. F. Rosenberger: *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley, 2006.
- O. Kempthorne: The randomization theory of experimental inference. *JASA* **50** (1955), 946–967.
- O. Kempthorne: Why randomize? *JSPI* **1** (1977), 1–25.
- O. Kempthorne, G. Zyskind, S. Addelman, T. N. Throckmorton and R. F. White: *Analysis of Variance Procedures*. Aeronautical Research Lab, 1961.
- J. A. Nelder: The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Roy. Soc. A* **283** (1965), 147–162.
- V. Plamadeala and W. F. Roseberger: Sequential monitoring with conditional randomization tests. *Annals of Statistics* **40** (2012), 30–44.

References IV

- S. J. Pocock and R. Simon: Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* **31** (1975) 103–115.
- F. Yates: The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments. *Biometrika* **440** (1939), 440–466.
- F. Yates: Contribution to discussion of 'The validity of comparative experiments' by F. J. Anscombe. *JRSSA* **111** (1948), 204–205.
- S. T. Ziliak: W. S. Gosset and some neglected concepts in experimental statistics: Guinessometrics II. *J. Wine Economics* **6** (2011), 252–277.