

**Introduction to Randomisation**

This research is supported in part by the Teaching and Learning Research Initiative (www.tlri.org.nz)

Software is available from the iNZight website ([www.stat.auckland.ac.nz/~wild/iNZight/](http://www.stat.auckland.ac.nz/~wild/iNZight/))

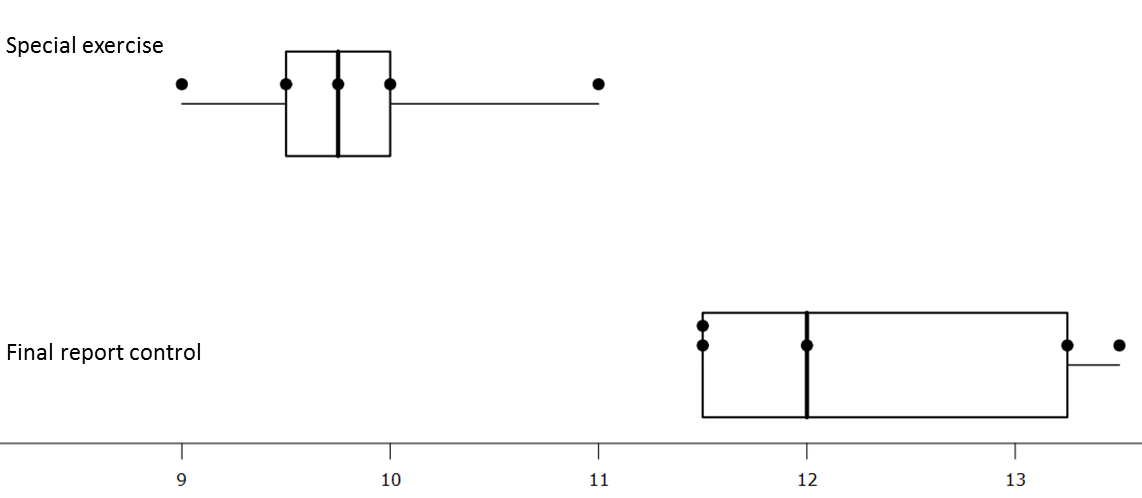
Resources are available on the CensusAtSchool website ([www.censusatschool.org.nz](http://www.censusatschool.org.nz))

**Randomisation Tests**

**The Walking Babies Experiment[[1]](#footnote-1): Does a special exercise programme lower walking age?**

All participants in this study were volunteers. 10 male infants (& parents) were *randomly assigned* to either the exercise group or the control group. The ages (in months) when these infants first walked without support are shown below.

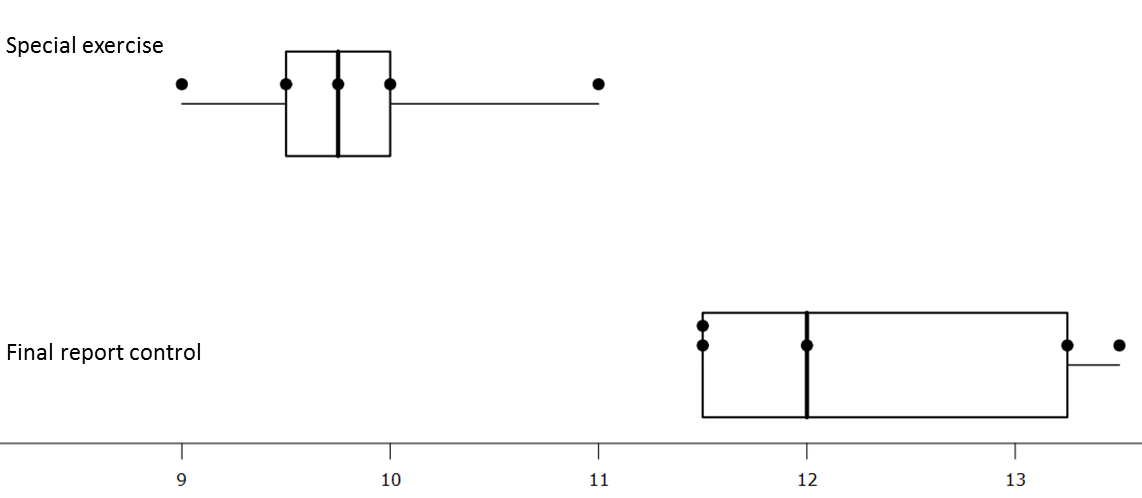
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Treatment** | **Age (months)** | | | | |
| Exercise | 9 | 9.5 | 9.75 | 10 | 11 |
| Control | 13.25 | 11.5 | 12 | 13.5 | 11.5 |



Control

Exercise

Age (months)



* Does it appear that these data provide evidence that the exercise programme is effective? **Yes**
* Is it possible that these babies’ walking ages have nothing to do with whether they undertook the exercises or not, (i.e*., it doesn’t matter which group they were randomly assigned to, they would still have the same walking age*) and the difference pattern we are observing is purely and simply the result of the luck-of-the-draw as to which babies *just happened by chance* to be assigned to which group and nothing else? **Yes**  
   Under this scenario we say: **“Chance is acting alone”.**

**Possible explanation:**

One *possible* explanation for the observed difference between these two groups:

**Chance is acting alone** (the exercise has no effect)

**Is the ‘*chance alone’* explanation simply *not plausible*?**

* Would our observed difference be *unlikely* when chance is acting alone?
* *How* do we determine whether an observed difference is *unlikely* when chance is acting alone?

Answer:   
*See* what’s **likely** and what’s **unlikely** *when chance is acting alone.*

(Reminder: *When chance is acting alone,* it doesn’t matter to which group the infants are randomly assigned,  
 they will still have the *same* walking age.)

9

11

10

12

14

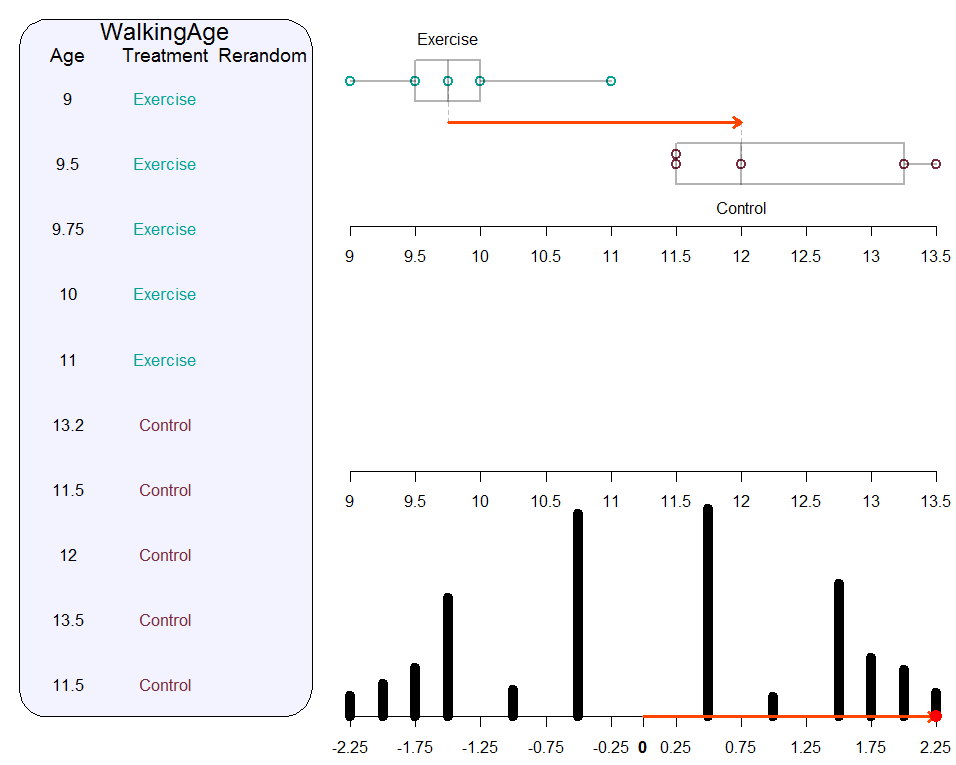
13

Exercise

Control

months

**Observed data**



months

months

months

**Chance alone**

Difference = Control median – Exercise median

=

9

11

10

12

14

13

Exercise

Control

9

11

10

12

14

13

Exercise

Control

9

11

10

12

14

13

Exercise

Control

**Re-randomisation 1**

**Re-randomisation 2**

**Re-randomisation 3**

Difference = Control median – Exercise median

=

Difference = Control median – Exercise median

=

Our **observed difference**

= Control median – Exercise median

= 12 – 9.75 = **2.25 months**

months

-3

-2

0

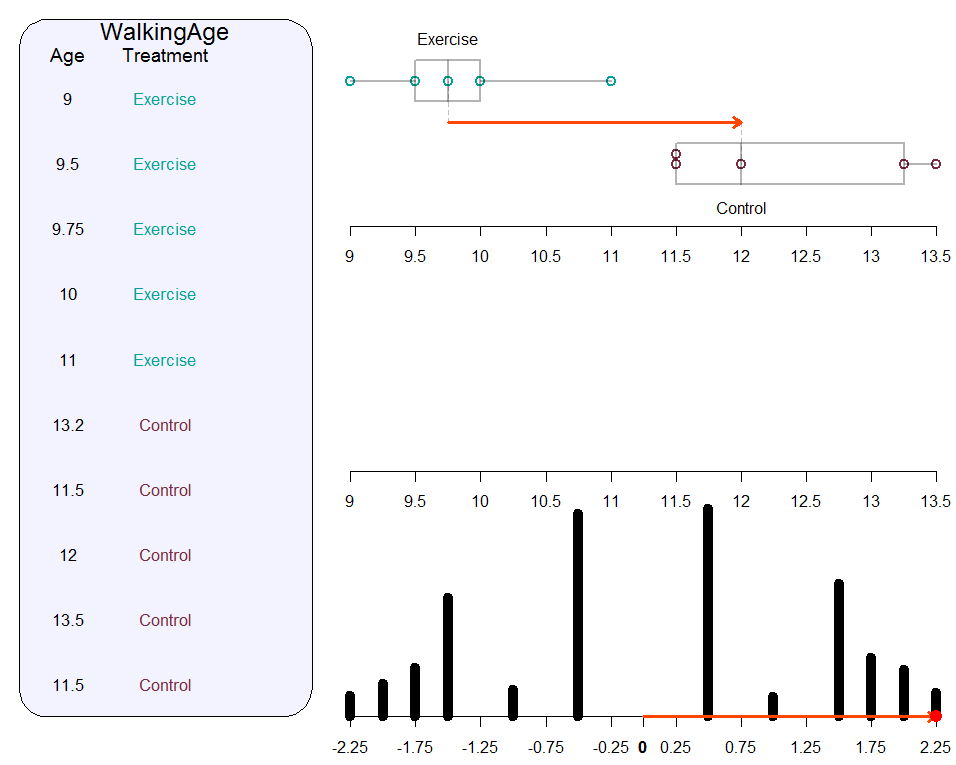
2

1

3

-1

**Re-randomisation distribution of differences  
 (under *chance alone*)**



Tail proportion:   
**27/1000 = 2.7%**

Is *chance alone* likely to generate differences as big as *our difference*?

**Hardly ever, they’re almost always smaller under chance alone.**

**.**

Our observed difference

= 2.25 months

Re-randomisation distribution of 1000 differences   
(under *chance alone*)

Our tail proportion of **2.7%** means:

* **27 times***out-of-a-1000 times we get a difference between medians of 2.25 months or more, when chance is acting alone. (This shows us that ‘chance is acting alone’ is a possible explanation but . . .)*
* *Under chance alone, it would be* **highly unlikely**  *to get a difference equal to or bigger than our observed difference of 2.25 months.*
* *An observed difference of 2.25 months or greater is highly unlikely when chance is acting alone . . .*  therefore, it probably is NOT acting alone.

**It’s a fairly safe bet chance is NOT acting alone.**

* We **will rule out** ‘chance is acting alone’ as a **plausible** explanation for the observed difference between the two groups.
* We have evidence**against** the ‘chance is acting alone’ explanation!
* We have evidence that **chance is not acting alone.**

**Conclusion:**

(If chance is not acting alone, then what else is also acting to help produce this observed difference? Remember: random assignment to 2 groups and each group receives a different treatment.)

Because these male infants (& parents) were **randomly assigned** to the groups then we may **claim** that the  **exercise was effective** in lowering the walking age. (Random assignment allows us to classify all explanations other than the difference-in-treatment explanation as ‘chance explanations’.)

Because the subjects were volunteers ( **not randomly selected** ), then we would need to consider carefully as to which wider group(s) this conclusion may apply.

**The logic behind the ‘chance alone’ argument:**

|  |  |  |
| --- | --- | --- |
|  | **Did she brush her teeth?** | **Is the exercise programme effective?** |
| **1.** **Statement to test** | She has brushed her teeth. | **Chance is acting alone.**  (The exercise has no effect.) |
| **2. Data** | The toothbrush is dry | **Difference between medians = 2.25 months** |
| **3.** **Consider 1. and data**  *If 1. is true, then what   are the chances of   getting data like that in 2.?* | ‘The toothbrush is dry’ would be highly unlikely, if she had brushed her teeth. | A difference of  **2.25 months or greater is highly unlikely**  when chance is acting alone.(Tail prop = **2.7%** ) |
| **4. Review statement 1**.   in light of 3. together   with data in 2.  (Is statement 1. plausible?) | Therefore, she probably has not brushed her teeth.  We have evidence that she has not brushed her teeth. | Therefore, it’s a fairly safe betchance is **not acting alone.**  We have evidence against‘chance is **acting alone’.**  We have evidence that something other than just ‘chance’ is acting. |

**Do the actual exercises lower the walking age?**

In the actual experiment there was another treatment group called the Control exercise group. The Control exercise group was treated with exactly the same conditions as the exercise group, i.e., the babies were exercised 12 minutes per day, except the parents were not told what specific exercises to perform. The purpose of this control exercise group was to help answer the question: If the special exercises are effective, then is it the actual nature of the exercises themselves which accounts for the difference or is it something else, e.g., simply the 12 minutes of daily physical contact with the baby? The following example is adapted.

10 male infants (& parents) were *randomly assigned* to either the exercise group or the control exercise group. The ages (months) when these infants first walked without support are shown below. (We have chosen to use group means rather than group medians in this example just to demonstrate that the randomisation method can be applied to means as well as medians.)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Treatment** | **Age (months)**  Our **observed difference** between the group means: | | | | |
| Exercise | 9 | 9.5 | 9.75 | 10 | 13  **= 1.4** **months** |
| Control exercise | 11 | 10 | 15 | 11.75 | 10.5 |

9

11

10

12

14

13

15

Exercise

Control.ex

Age (months)

**Possible explanation**

One *possible* explanation for the observed difference between these two groups:

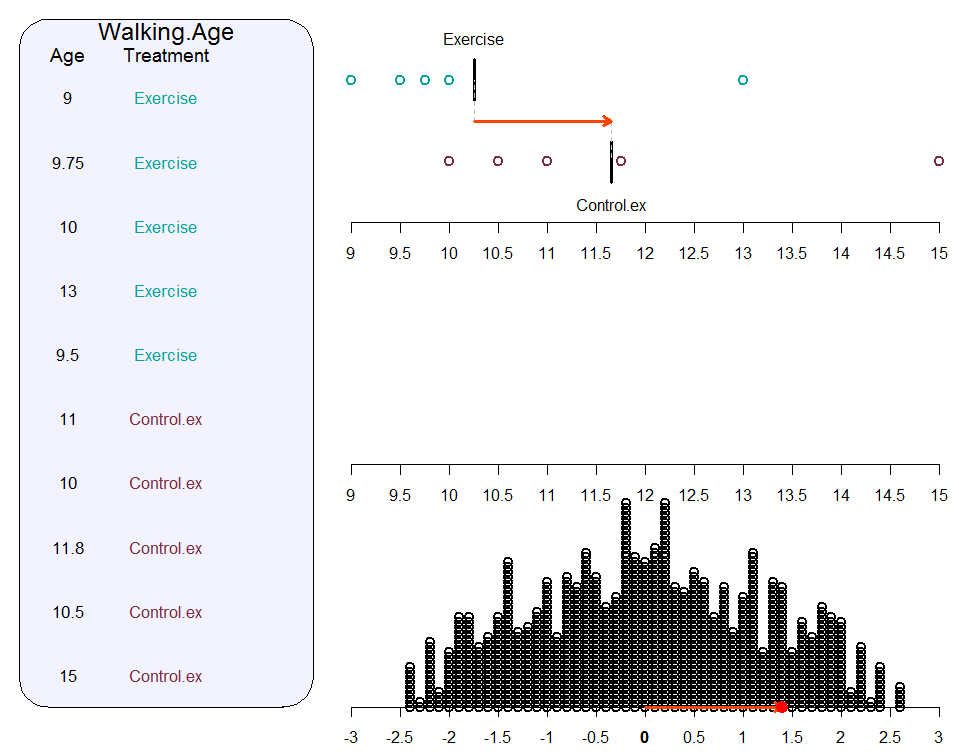
**Chance is acting alone** (the exercise has no effect)

**Is the ‘*chance alone’* explanation simply *not plausible*?**

* Would our observed difference be *unlikely* when chance is acting alone?
* *How* do we determine whether an observed difference is *unlikely* when chance is acting alone?

Answer:   
*See* what’s **likely** and what’s **unlikely** *when chance is acting alone.*

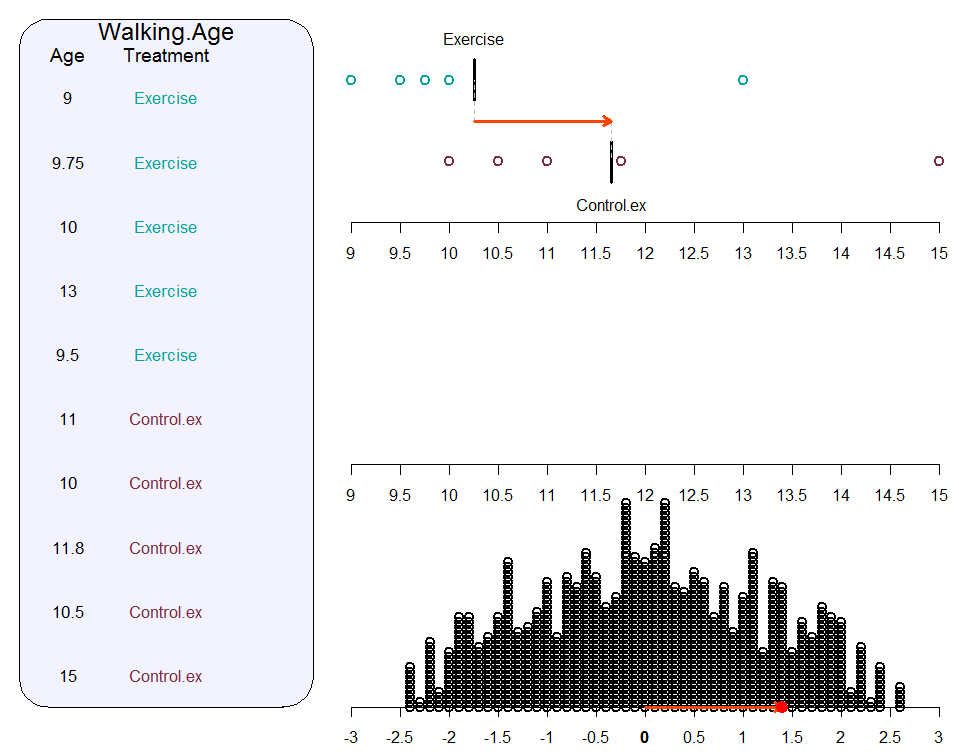
(Reminder: *When chance is acting alone,* it doesn’t matter to which group the infants are randomly assigned,  
 they will still have the *same* walking age.)



**It’s not unusual to get differences this big or bigger under chance alone.**

Our observed difference  
 = 1.4 months

Is *chance alone* likely to generate differences as big as *our difference*?



Tail proportion:   
**150/1000 = 15%**

Re-randomisation distribution of 1000 differences   
(under *chance alone*)

Our tail proportion of  **15%** means:

* **150 times***out-of-a-1000 times we get a difference between means of 1.4 months or more, when chance is acting alone.*
* *Under chance alone, it’s not unusual**to get a difference equal to or bigger than our observed difference of 1.4 months.*
* Therefore, **chance could be acting alone.**
* We **can** **NOT rule out** ‘chance is acting alone’ as a **plausible** explanation for the observed difference between the two groups.
* We have **no**  evidence**against** the ‘chance-is-acting-alone’ explanation!
* Chance COULD be acting alone BUT something else, as well as chance, **COULD** ALSO be acting.

**Conclusion**

(If chance is not acting alone, then what else could also be acting to help produce this observed difference? Remember: Random assignment to 2 groups and each group receives a different treatment.)

Because the male infants (& parents) were **randomly assigned** to the groups, we conclude that the observed difference is the result of:

*EITHER* **chance acting alone**    
 *OR* **an exercise effect together with chance acting.** –we do *NOT* have *ENOUGH INFORMATION* to *MAKE A CALL* as to which.

**The logic behind the ‘chance alone’ argument:**

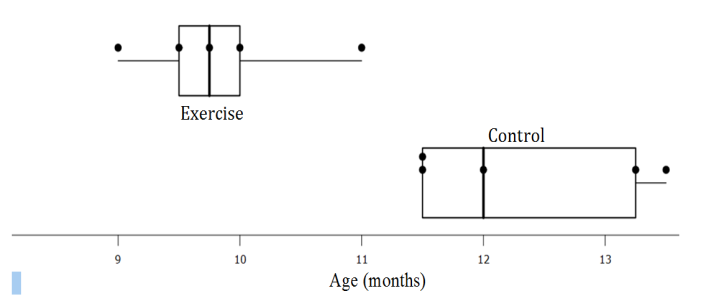
|  |  |  |
| --- | --- | --- |
|  | **Did she brush her  teeth (2)?** | **Is the *actual* exercise effective?** |
| **1.** **Statement to test** | She has brushed her teeth. | Chance is acting alone.  (The exercise has no effect.) |
| **2. Data** | The toothbrush is wet | Observed difference between means = 1.4 months |
| **3.** **Consider 1. and data**  *If 1. is true, then what   are the chances of   getting data like that in 2.?* | ‘The toothbrush is wet’ would be likely if she had brushed her teeth. | A difference of 1.4 months or greateris **not unusual when chance is acting alone.**  (Tail prop = **15%**  ) |
| **4. Review statement 1**.   in light of 3. together   with data in 2.  (Is statement 1. plausible?) | Therefore, she could have brushed her teeth (or she could have run the brush under the tap).  We have no evidence that she has not brushed her teeth. | Therefore, chance could be acting alone OR **something else could be acting along with chance.**  We don’t have enough information to make the call as to which one. We can’t claim that the actual exercise is effective. |

**Randomisation Tests**

**The Big Ideas:**

Can we rule out the ‘chance alone’ explanation because it’s implausible?

1. Looking at the world using data is like looking through a window with ripples in the glass:

*‘What we see is not quite the way it really is’*

1. The **possibility** that **chance acting alone** has produced the observed difference pattern
2. **Assessing** the plausibility of the ‘chance alone’ explanation
3. **Forming a conclusion** about what the data are telling us

**What do we mean by *chance alone*?**

***Under chance******alone*** the difference pattern we observe is purely and simply the result of the luck-of-the-draw as to which units (e.g., babies) *just happened by chance* to be assigned to which group and nothing else.   
Under chance alone, the value of the measure made on a unit (e.g. a baby’s walking age) has nothing to do with the group to which that unit (baby) had been assigned, i.e., **it doesn’t matter to which group the units** (babies) **were randomly assigned, each unit** (baby) **would still have the same actual value of the measure made** (walking age)**.**

**How do we assess the plausibility of the chance alone explanation?**

**Randomisation Test** ( 3Rs: ***R****e-randomise,* ***R****epeat,* ***R****eview* )

**Re-randomise:**

Ignore each unit’s (baby’s) assigned group, randomly re-assign each unit to a group – always keeping its actual measure, that is, **re-assign as if chance is acting alone**. Determine and record the value of the statistic of interest (e.g., ).

**Repeat:**

Repeat the re-randomisation a large number of times. Plot the recorded values of the statistic of interest to generate the re-randomisation distribution under chance alone. See what are likely and what are unlikely (hard to get) values for the statistic of interest (e.g., ) when chance is acting alone.

**Review:**

Review the ‘chance-is-acting-alone’ explanation by locating the observed difference (e.g., the actual observed data value of ) on the re-randomisation distribution.

* An observed difference located far out in the tails discredits the chance-alone explanation. The observed difference would be highly unlikely if chance were acting alone, therefore chance probably isn’t acting alone. It gives us: evidence that chance is not acting alone; evidence that something other than just chance is acting.
* An observed difference not located far out in the tails does not discredit the ‘chance-alone’ explanation. The observed difference would be not be surprising if chance were acting alone, therefore chance *could* be acting alone. But equally, something as well as ‘chance’ *could* also be acting.

**Guide Lines for Assessing ‘Chance alone’**

When the **tail proportion** in the re-randomisation distribution is **less than 10%** then:

* the observed difference would be unlikely when chance is acting alone, therefore it’s a fairly safe bet chance isn’t acting alone.
* we have evidence **against** ‘chance is acting alone’
* we have evidence that **chance is not acting alone**

When the **tail proportion** in the re-randomisation distribution is bigger than **about 10%** then:

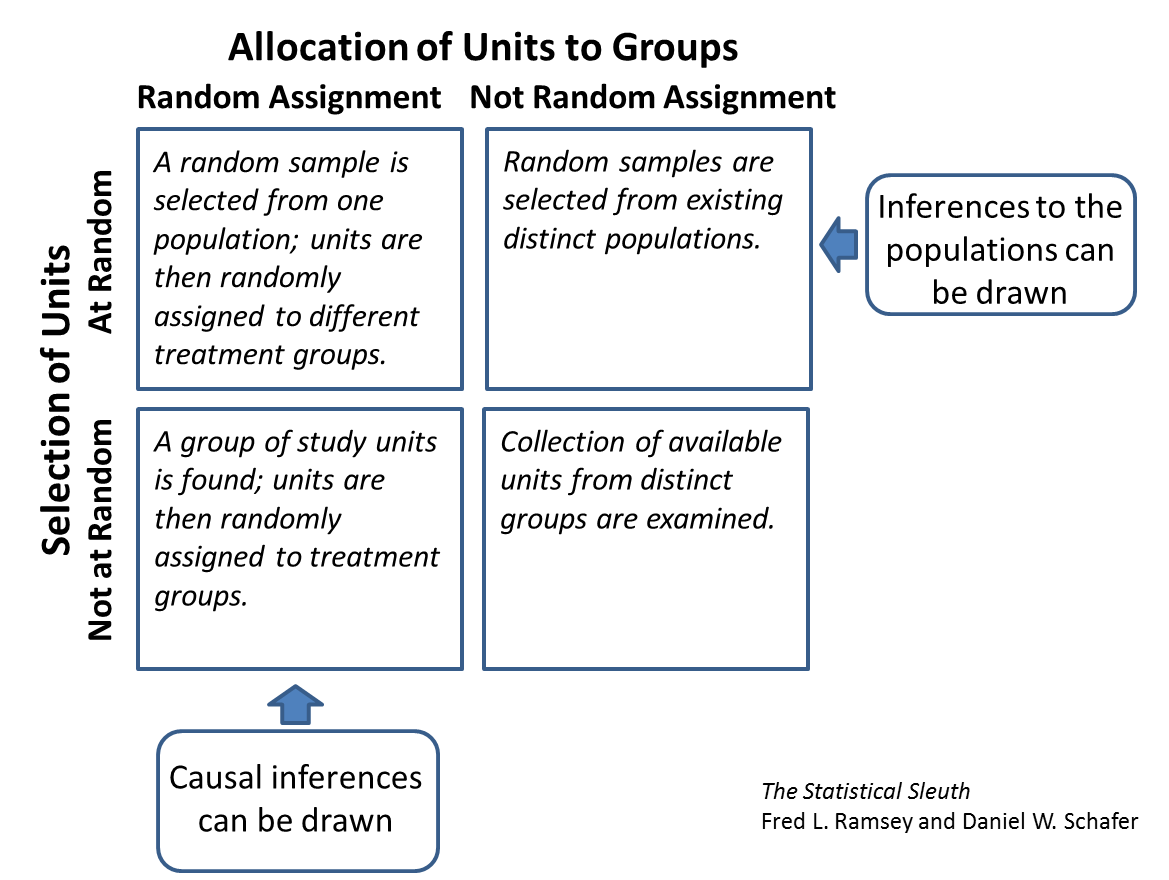
* the observed difference is not unusual when chance is acting alone, therefore chance **COULD** be acting alone
* we have **NO** evidence **against** ‘chance is acting alone’
* ‘chance’ COULD be acting alone ***OR*** something else as well as ‘chance’ COULDalsobe acting (we don’t have enough information to determine which one of these two possibilities applies).

**The Link between the Randomisation** (data production) **and the Conclusion**

**Randomisation (data production)**

1. Random assignment of units to treatment groups (an experiment)
2. Random sampling of units from a population or populations

**Where is the randomisation and what inference does it support?**



1. Adapted from: Phillip R. Zelazo, Nancy Ann Zelazo, & Sarah Kolb, ‘*Walking’ in the Newborn,* Science, Vol. 176 (1972), pp314-315 and *Statistics from Data to Decision*: Watkins, Scheaffer, & Cobb page 515 [↑](#footnote-ref-1)