Tutorial: designing and interpreting controlled experiments. "Why do scientists do what scientists do?"

Working in groups, read through the following scenario's. Some are based on real events where as some are fictitious. The goal is to gain an understanding of what data really means. After a 90 min, we will discuss the different ideas generated by the groups.

Positive and negative controls

β-adrenoceptors stimulate adenylate cyclase to generate cAMP in airway smooth muscle which leads to muscle relaxation. An experimental β-adrenoceptor agonist called "keeloline" was tested for its ability to induce accumulation of cAMP in airway smooth muscle cells. 5 different types of experiment were conducted in the order 1-5, each with 3 different replicates and the results are shown in the table below

Experiment		*cAMP (nmol ml^{-1})*
1	No drug	21, 22, 23
2	Keeloline	27, 26, 27
3	Inactive Keeloline analogue	27, 27, 28
4	Salbutamol	379, 385,342
5	Alternative salbutamol result	29, 30, 27

Look at the table and consider the following scenarios one at a time

a) Experiment 1 shows the cAMP measured after no drug treatment, whereas experiment 2 shows the cAMP after treatment with a high dose of keeloline (you can assume it causes the drug's maximum effect). Based on the changes in cAMP, do you think that keeloline is a β-adrenoceptor agonist ? (Remember at this stage you only have results from expts 1 and 2)
b) Experiment 3 shows cAMP levels after addition of a drug that is an analogue of keeloline but which is known not to bind β-adrenoceptor (a "negative control"). Does this affect your previous conclusion?
c) Experiment 4 shows cAMP accumulation induced by salbutamol, a known β-adrenoceptor agonist (a "positive control"). Does this affect your previous conclusion?
d) What would you conclude if the response to salbutamol that was observed in experiment 4 had not been obtained, but instead salbutamol had the effect shown in experiment 5? (i.e. pretend expt 4 doesn't exist)

The question for you to consider is this: how have "control" experiments with drugs which have a known activity altered your perception of the novel drug?

Fair comparisions

1. Imagine a drugs that bind to a particular receptor and which are expected to kill cancer cells. You wish to add the drug to the cells and determine how many cells die. But suppose the drug is insoluble in aqueous media at high concentrations, so you decide to dissolve it first in DMSO (an organic solvent) and then dilute it in an aqueous solution before adding it to the cells.
    a. How will you measure the effect of the drug? The cells treated with the drug have also been exposed to DMSO, but the cell that haven't been treated with drug have not been exposed to DMSO. So are you measuring the effect of the drug or the DMSO ? How can you make this comparison fair?
    b. What would you do if you wanted to measure different concentrations of drug? Wouldn't the concentration of DMSO change too as you dilute the drug? How can you make this comparison fair?
    c. If instead of one drug you wish to compare two drugs but one drug is soluble in ethanol and the other in DMSO. You make a solution of one drug in ethanol and the other in DMSO, dilute them both in aqueous and then add each drug to the cells. Is it reasonable to compare the results of the two drugs, because with one drug, the cells will also be exposed to DMSO and in for the other drug the cells will also be exposed to ethanol? What could you do to make the comparison fair?
    d. Supplementary question. Without going into too much detail, which experiments could you do to prove the drug is acting through the receptor as expected? (It could kill cells by acting through another receptor of which you are unaware).

2. The slides below show the characteristics of patients entering the ICON7 trial to evaluate the efficacy of bevacuzimab in ovarian cancer. The slides were shown at a conference (Prague, ICGS 25 Oct 2010) where the results were reported. The efficacy of bevacuzimab has been evaluated by adding it to the current standard of care, which is surgery followed by chemotherapy comprising carboplatin and paclitaxel. The trial compared chemotherapy (labelled TC) versus chemotherapy plus bevacizumab (labelled TCBev). The patients were allocated to receive either TC or TCBev. What was the point the speaker was trying to make?

KEY TO ABBREVIATIONS
ECOG PS = performance status, a measure of how healthy the patients were.
Histology- a measure of different types of ovarian cancer
Grade= tumour grade, a measure of how aggressive it is
FIGO stage = measure of how far the cancer has spread
Debulking surgery residual tumour- a measure of how much of the tumour remains after surgery to remove the tumour

Baseline characteristics

Characteristic	TC (n=764)	TCBev (n=764)
Median age (range)	57 (18–81)	57 (24–82)
ECOG PS, n (%)		
0	358 (47)	334 (45)
1	354 (47)	366 (49)
2	43 (6)	45 (6)
Origin of cancer, n (%)		
Ovary	667 (87)	673 (88)
Fallopian tube	29 (4)	27 (4)
Primary peritoneal	56 (7)	50 (6)
Multiple sites	12 (2)	14 (2)
Histology		
Serous	529 (69)	525 (69)
Clear cell	60 (8)	67 (9)
Endometrioid	57 (7)	60 (8)
Mucinous	15 (2)	19 (2)
Mixed/other	103 (13)	93 (12)
Grade, n (%)		
1	56 (7)	41 (5)
2	142 (19)	175 (23)
3	556 (74)	538 (71)
Unknown	10	10

Based on presentations by T. Pujol, A. Oza et al. at ESMO 2010 and J. Pfisterer et al. at IGCS 2010

Baseline characteristics (cont'd)

Characteristic, n (%)	TC (n=764)	TCBev (n=764)
FIGO stage, n (%)		
I/IIA	75 (10)	67 (9)
IIB–IIIB	160 (21)	155 (20)
IIIC/IV	529 (69)	542 (71)
Debulking surgery		
residual tumor ≤1 cm	552 (74)	559 (74)
residual tumor >1 cm	195 (26)	192 (26)
No surgery	17 (2)	13 (2)
FIGO stage and residuum*		
Stage I–III (≤1 cm)	508 (66)	518 (68)
Stage I–III (>1 cm)	150 (20)	140 (18)
Stage III (inoperable)/IV	106 (14)	106 (14)
Intent to start chemotherapy*		
≤4 weeks from surgery	328 (43)	326 (43)
>4 weeks from surgery	436 (57)	438 (57)

Based on presentations by T. Pujol, A. Oza et al. at ESMO 2010 and J. Pfisterer et al. at IGCS 2010

Evidence of absence and absence of evidence

1. Its relatively easy to prove that a drug has a particular pharmacological activity (eg killing cancer cells) because you can simply measure the response to the drug. But it is more challenging to demonstrate that a drug lacks a particular activity. For example, a vaccination is currently being introduced across the UK to prevent cervical cancer. How would your attempt to **prove** that a particular vaccine is safe (doesn't have unwanted side effects)?

2. Its quite common to may hear someone being interviewed on the television news saying something similar to:

    "There is no evidence linking this drug with suicidal thoughts"

What does this mean?

(Suicidal thoughts are simply an example here, it could be any adverse event associated with a drug. A number of drugs have been withdrawn because of this adverse event so it is a very relevant example, and this has been discussed in high profile cases in the press. You may also have to explain to a patient why there was a warning on tv about a drug he/she is taking so this is a very relevant question!.)

Normalization

1. Imagine someone proposes a theory that eating crisps makes children misbehave. To test this, you measure crips consumption by counting the number of crisp packets in the rubbish bin at the end of the day in Mr Maddock's class (which is well behaved) and in Miss Curtis's class (which is badly behaved). The results are shown below.

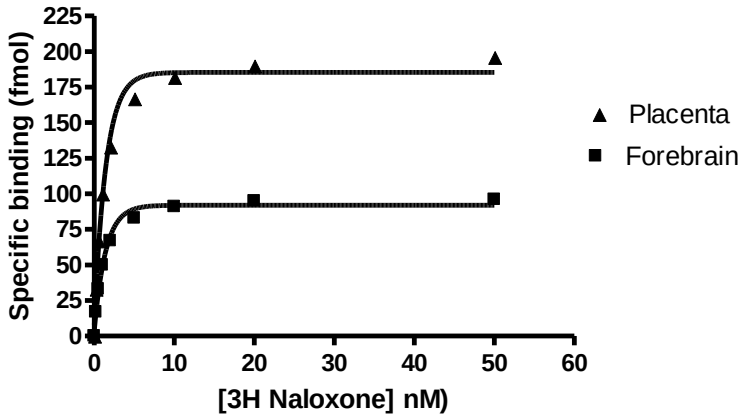
Teacher	**Mr Maddock**	**Miss Curtis**
Behaviour	Well behaved	Badly behaved
Number of crisp packets	22	31

What do you conclude ? Now read the note on the last page and see if your conclusion changes.

2. The table and graph below shows specific binding of the opiate antagonist to cell membranes prepared from forebrain and from placenta. The experiments were conducted using 1 ml of the membrane preparation. The easiest way to quantify the concentration of cell membranes in each preparation is to measure the membrane protein content. Does placenta or forebrain contain more receptor?

The membrane protein content for the samples used in the binding assays was:
Forebrain 2 mg/ml
Placenta 10 mg/ml

	Forebrain	**Placenta**
[^{3}H Naloxone]	**Specific binding (fmol)**	**Specific binding (fmol)**
0	**0**	**0**
0.2	**17**	**33**
0.5	**33**	**67**
1	**50**	**100**
2	**67**	**133**
5	**83**	**167**
10	**91**	**182**
20	**95**	**190**
50	**96**	**196**

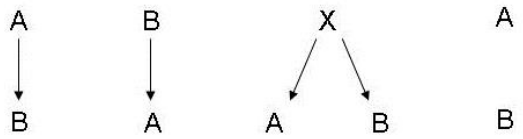


Cause and correlation

Some interesting correlations to think about first:

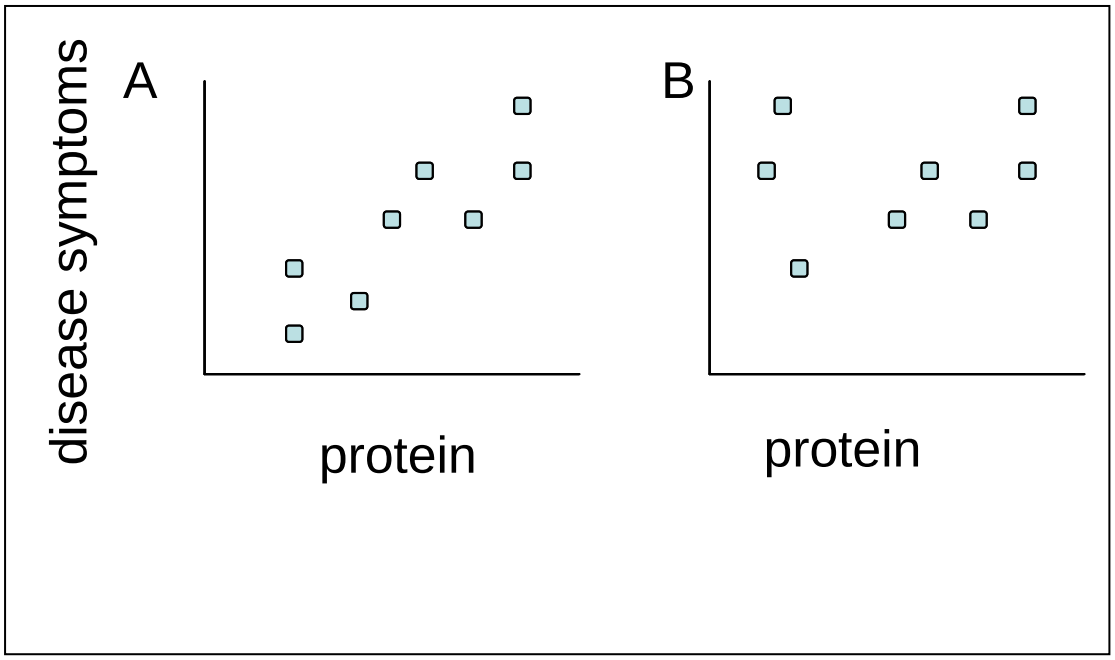
1. There is a clear link between marriage and divorce: people who are married are far more likely to divorce than those who are not married
2. Ice cream sales and the number of shark attacks on swimmers are correlated
3. It is claimed that skirt lengths and stock prices are highly correlated
4. There is a strong correlation between the number of tooth fillings in school children and their vocabulary size

Try to come up with some explanations. The following diagram might help



5. Try this website http://www.tylervigen.com/ for more correlations (don't spend too long!)

6. Alzheimer's disease



The diagrams shows the relationship between the abundance of two proteins (A and B) in brains of patients suffering from Alzheimer's disease.

a) Is there any link between the expression of these proteins and the disease ?
b) Does accumulation of either protein cause Alzheimer's disease?

c) How would you test these experimentally? (We aren't looking for too much detail here, just general ideas)

7. What do think of the following correlation (in the graph)?

The Correlation Among Obesity, Apnea-Hypopnea Index, and Tonsil Size in Children*

Yuen-yu Lam, FHKAM(Paed); Eric Y.T. Chan, MRCPCH; Daniel K. Ng, FHKAM(Paed); Chung-hong Chan, BSc; Josephine M.Y. Cheung, RPSGT; Shuk-yu Leung; Pok-yu Chow, FHKAM(Paed); and Ka-ki Kwok, FHKAM(Paed)

Background: **The correlation between obesity and severity of obstructive sleep apnea (OSA) is well established in adults, but data are inconsistent in children. We hypothesized that there is a significant correlation between the degree of obesity and the severity of OSA in children.**
Methods: **We retrospectively reviewed records of weight, height, history, and polysomnography of all 1- to 15- year-old children referred to our sleep laboratory. Children with known anomalies and repeated polysomnography were excluded from this study. Obesity was defined as body mass index z score (BMI Z score) > 1.96. The correlation between BMI Z score and apnea-hypopnea index (AHI) was assessed. Possible confounding factors, *ie*, age, gender, and tonsil size, were adjusted by multiple linear regression.**
Results: **Four hundred eighty-two children were included in this study. Obese children had a significantly higher AHI (median, 1.5; interquartile range [IQR], 0.2 to 7.0) than the AHI of nonobese children (median, 0.7; IQR, 0.0 to 2.5). BMI Z score was significantly correlated with log-transformed AHI (Ln[AHI]) [$r = 0.156$, $p = 0.003$]. BMI Z score and tonsil size were still correlated with Ln(AHI) even after adjusted for other confounding factors ($p = 0.001$).**
Conclusion: **Degree of obesity as measured by BMI Z score and tonsil size are significantly related to severity of OSA as reflected by the AHI, although the correlation is mild.**
(CHEST 2006; 130:1751–1756)

Key words: child; obesity; obstructive; polysomnography; sleep apnea

Abbreviations: AHI = apnea-hypopnea index; BMI = body mass index; BMI Z score = body mass index z score; IQR = interquartile range; Ln(age) = log-transformed age; Ln(AHI) = log-transformed apnea-hypopnea index; OSA = obstructive sleep apnea; OSAS = obstructive sleep apnea syndrome; SDB = sleep-disordered breathing

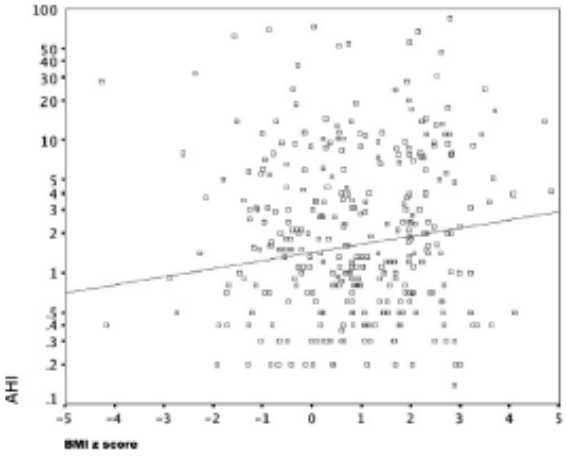


FIGURE 1. Correlation between AHI and BMI Z score (Y axis is in log scale).

What does the data really tell you?
"Over- interpretation" of data -What (you think) you see is (not necessarily) what you get.

Sometimes, people present data and "over-interpret it". For example, if its sunny the first week of May and someone concludes we are about to have a wonderful summer, its probably a bit premature. Consider the following example of FLUX through a pathway

Imagine the following biochemical scheme involving3 3 metabolites, A,B and C.

A →B→C

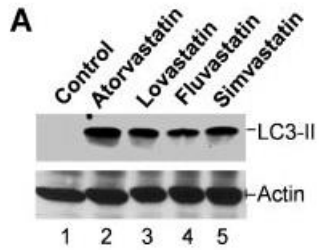
Now suppose you do something to stimulate a cell, and you measure decrease in the amount of B. What is your conclusion? Is it that you have inhibited A →B or could it be that B→C happens more quickly?

If, on the other hand, you stimulate the cell in a different way and you detect an increase in B is that because A →B occurs faster or B→C occurs more slowly ?

An example of this from healthcare - if you see an increase in the incidence of a disease being diagnosed, why is this?

- Increased incidence of the disease ?
- Improved efficiency of detecting the disease?
- Both?

Another example, this time involving statins (drugs used to reduce cholesterol). Autophagy is a process in which cells catabolize their own organelles. Autophagy requires the accumulation of a protein called LC3-II. Statins were proposed to induce autophagy in cells (this probably isn't involved in the use of statins in regulating cholesterol levels). The data that led to this was obtained by treating cells with the statin, then measuring the increase in LC3-II by western blotting. Here is some data showing the accumulation of LC3-II, measured by western blottinG, after treating cells with various statin (from The Prostate, 2010, 70, 971).



(Control represents addition of the solvent that the drugs were dissolved in)
At first glance, you could conclude that statins stimulate the production of LC3-II and so this demonstrates that statins induce autophagy. But you could also argue that the statins inhibit the degradation of LC3-II, and because there is continual synthesis of small amounts of LC3-II even in cells not undergoing autophagy, this leads to the

accumulation of LC3-II. In this scenario, where the statin is proposed to inhibit the degradation of LC3-II, the statin would be inhibiting autophagy. So the same data may be interpreted in two completely different ways! (additional experiments were needed to resolve the issue.)

A conversation between two astonomers in the 19th century (this example is borrowed from Carl Sagan's book "Cosmos").

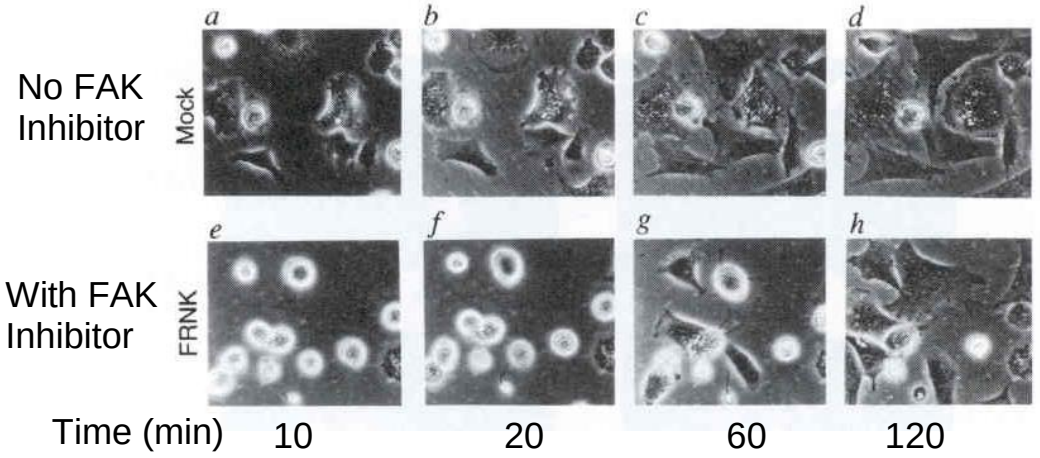
"There is absolutely nothing to see on Venus when I look with my telescope"
"I wonder why that is? Maybe the surface is obscured by clouds"
"That's possible. What type of landscapes are covered by clouds?"
"Well, landscapes with lots of liquid on the surface"
"That's right. What types of planets have lots of liquids on their surface?"
"Oh, I don't know. Planets with seas, I guess, and maybe marshes."
"Good point. What do we find in the sea or in marshland ?"
"Well, often there are fish, or seabirds, maybe some mammals that like water."

Observation: there is nothing to see on Venus
Conclusion: Venus is teaming with wildlife.

Think back to the example on LC3-II and autophagy. Does the data tell you whether statins have any effect on autophagy (irrespective of whether its and increase or decrease)?
In fact, all the data tells you is that statins alter the levels of LC3-II. Autophagy wasn't really measured directly.
Suppose a new statin is developed and is shown in clinical trials to reduce significantly plasma cholesterol levels. It would be tempting to say that the new statin reduces mortality from heart attacks caused by arterial plaques. But unless the clinical trial actually measured patient's lifespan, it wouldn't be safe to make that claim. It might be probable, but all you could say with certainty is that cholesterol levels were lower. (Actually, not even that is completely certain….read the statistics example to find out why.)

Bias

1. An enzyme called focal adhesion kinase (FAK) was thought to regulate the ability of cells to spread out on extracellular matrix. This process can be visualized with a microscope by taking cells in suspension, where they are essentially spheres, allowing them to attach to a flat plastic dish and taking photographs of them after 20 minutes, to see how much they have spread (flattened out – see diagram for an example). To assess the role of focal adhesion kinase in this, some of the cells were treated with an inhibitor of FAK whereas others were treated with an appropriate control (and no FAK inhibitor). The cells were then added to the extracellular matrix and photos were taken of the cells over 2 hours. To assess the effect of inhibiting FAK on delaying cell spreading, the number of cells that were spread after 20 minutes was counted in each sample. The entire experimental procedure was carried out by one scientist.



Where is the obvious source of bias, and how can you eliminate it? (A salient point that you might not be aware of is that scientists are assessed by the number of papers they publish, and this depends on the number of novel discoveries they make.)

(Inhibitors of focal adhesion kinase are currently in clinical trials for the treatment of cancer. By inhibiting cell spreading they may inhibit metastasis)

2. Clinical trials of new drugs are often "placebo controlled". One group of patients receives the drug, while another group receives a preparation that lack the active ingredient. Sometimes the patients' response to the drug is hard to measure in an objective fashion (eg depression) and the assessment of the effectiveness of the drug may involve a doctor asking the patient a series of questions.

Where is there the potential for bias and how can this be avoided?

Error and statistics

1. Almost all pharmacological science involves pipetting at some point and researchers must become proficient at it. Imagine you are testing a new drug that has been proposed to be useful in the treatment of hypotension (low blood pressure). The drug may cause contraction of smooth muscle around capillaries, thereby increasing blood pressure.

To test this you pipette a range of different drug concentrations onto smooth muscle cells and measure the contraction to determine the potency (EC_{50}) of the drug. How will the results be affected if:

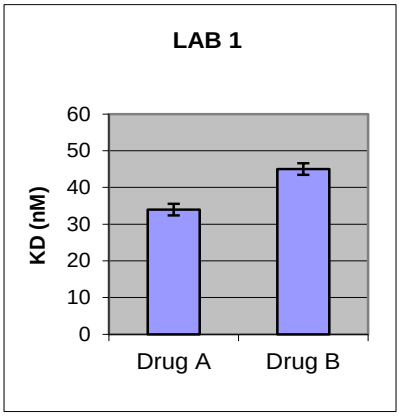
a) your pipette always pipettes 90% of the volume you expect it should?

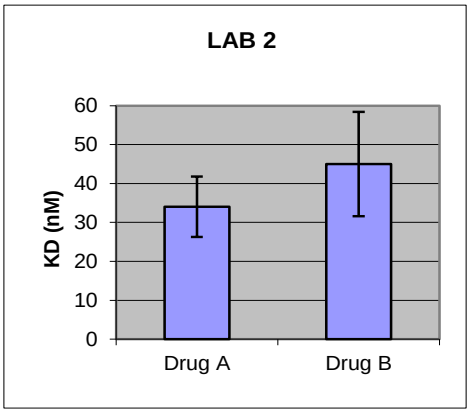
b) your pipette is worn and leaky, so each time you pipette the volume of drug you dispense varies randomly by 10%

2. Suppose you wish to compare to drugs, A and B, to see if one has a higher affinity than the other. Look at the following measurements of K_D made using binding assays. The table shows the individual measurement of K_D made in 5 different experiments. This has been repeated in two different laboratories (eg two different Universities)

	LAB 1	
Expt no	**Drug A**	**Drug B**
1	35	45
2	32	46
3	34	42
4	36	45
5	33	47
	Drug A	**Drug B**
Mean	34	45
S.D.	1.58	1.87
t-test		0.0010

	LAB 2	
Expt no	**Drug A**	**Drug B**
1	22	43
2	28	32
3	46	61
4	40	52
5	34	37
	Drug A	**Drug B**
Mean	34	45
S.D.	7.7	13.4
t-test		0.06

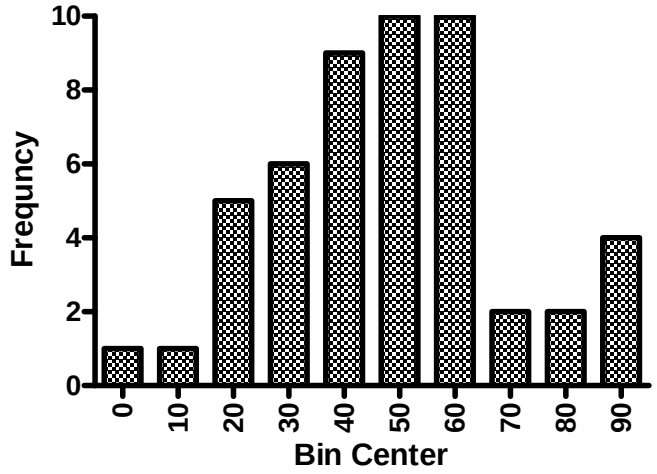




The graphs show the mean data, and the error bars on the graph show the standard deviations. Both laboratories have obtained the same mean results (K_D values of 34 and 45 nM for drugs A and B respectively). What would you conclude if you worked in LAB1. What would you conclude if you worked in LAB2 ?

3. Sampling error

A	B	A	B
1	55	26	38
2	75	27	2
3	51	28	33
4	20	29	26
5	70	30	22
6	33	31	53
7	38	32	41
8	55	33	61
9	47	34	43
10	87	35	91
11	61	36	41
12	50	37	42
13	32	38	59
14	50	39	46
15	55	40	48
16	72	41	28
17	64	42	45
18	61	43	77
19	39	44	6
20	59	45	21
21	89	46	48
22	23	47	48
23	25	48	89
24	39	49	44
25	55	50	15



The graph shows an analysis of the data shown in the table. The data have been put in "bins". For example, there are 5 occurrences of a number falling within the range 15-25 (bin centre 20).

This exercise is to illustrate sampling error. The data in the table listed in the columns labelled B have an average value of 47. The graph is provided simply so you have an idea of the scatter of the data .

Obtain 3 data points from column B as follows. Ask a colleague to call out 3 random numbers between 1 and 50 (without looking at the table). Then look down column A for these numbers and select the corresponding value listed in column B. Take the average of the values listed in column B. How does the result compare to the true average?

Repeat this exercise but this time try selecting just 1, 2 or 5 numbers. What do you conclude?

What do the numbers really mean?

1. It has been suggested that sunburn can double the risk of skin cancer. 9000 people (about 0.015% of the UK population) are year are diagnosed with melanoma each year in the UK. What is an individual's risk of melanoma if they get sunburn. Would a newspaper article with the headline which says "sunburn doubles the risk of skin cancer" be an accurate representation?

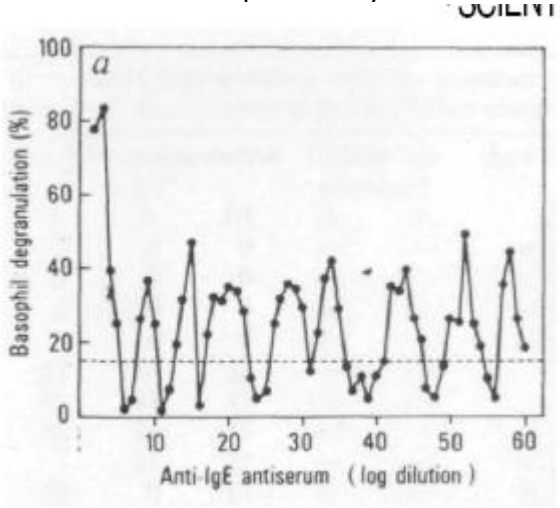
   (Don't be misguided by this example, advanced melanoma is a difficult disease to treat and is not trivial.)

   This type of analysis is also worth bearing in mind when thinking about the frequency of adverse reactions (side effects) to drugs.

2. A non-scientific example: next time you hear a politician claim that the rate of increase of inflation has fallen, what does that mean? Is inflation still going up or is it going down?
   (If you quite rightly think example this isn't directly relevant to pharmacy, simply replace "inflation" with "the cost of developing a new drug".)

A remarkable case of…….?

In 1988, Dr J Beneveniste and collegues was investigating the release of histamine from basophils triggered by an antibody. Understanding this is important in understanding the basis of allergy. It was found that the antibody (labelled anti IgE antiserum in the figure) could be diluted 10^{60} fold and it still caused histamine release. Furthermore, this activity appeared to "oscillate" with dilution. This was hard to understand because after such extensive dilutions, no antibody molecules would be expected to be left. This was interpreted by some as the water mimicking the structure of the antibody and it was seized upon as an explanation for the effectiveness of homeopathic remedies. The data was so provocative that it was published as an article in the premier scientific journal Nature, although there were also reservations expressed by the editors. What do you think is happening?



Note for Crisp packet exercise

There are 22 children in Mr Maddocks class and 31 in Miss Curtis's class.