Design of Experiment Exercise

Below are a series of published[1] or potential experiments. These are a selection from many different types of studies and are to illustrate that designing experiments correctly is a universal necessity and not a peculiarity of clinical spectroscopy.

You are to look at each of these patient metadata[2], and to consider the population demographics[3]. Discuss this in groups and decide if there are confounders that may affect the outcome of the study. Some Studies have prompts for you to consider.

Study 1

The table below is taken from Lawton, K.A. *et al*. (2008) *Pharmacogenomics* **9**, 383-397. The study involved metabolomics with GC-MS & LC-MS with the aim of discovering small molecules that could detect liver failure from plasma samples.

- *Do you think the study has been designed properly?*

Table 1. Demographic Information of the Healthy Group and Liver Failure Patient Group Investigated[a]

	healthy group ($n = 23$)	patient group ($n = 24$)
Gender (male/female)	15/8	21/3
HBsAg	Negative	Positive
Age (year)	27.39 ± 9.24	46.77 ± 13.35
ALT (U/L)	<40	172.63 ± 147.49
TB (μmol/L)	<12	457.33 ± 135.48
PT (s)	<14	26.06 ± 15.14
MELD score	/	24.68 ± 8.38

[a] Abbreviations: ALT, alanine aminotransferase; TB, total bilirubin; PT, prothrombin time; MELD, model for end-stage liver disease. The value is represented as the form of mean ± SD.

Study 2

An early studying using Affymetrix arrays with 6,817 DNA probes published in *Science* (Golub T.R. *et al*. (1999) *Science* **286**, 531-537) compared 27 individuals with acute lymphoblastic leukemia and people with 11 acute myeloid leukemia.

- *What might be wrong with this study design?*

---

[1] For the papers that are published in respected journals this exercise is not to suggest that any errors are deliberate.

[2] Metadata is a term that refers to the data about the data; *i.e.*, this provides information about the measurement data collected.

[3] These are the statistical data relating to the population and case-control groups.

Study 3

SELDI is a method that employs modified surfaces for MALDI thereby allowing for enrichment of certain proteins. The table below is taken from Navaglia *et al.* (2009) *Clinical Chemistry and Laboratory Medicine* **47**, 713-723.

- *There are many confounders in this study. How many can you find?*

Table 1 Clinical characteristics of the 4 subjects' groups included in the study.

	Pancreatic cancer	Chronic pancreatitis	Type 2 diabetes mellitus	Healthy controls
Total number of cases	126	61	24	12
Gender (cases)				
Males	72	47	12	2
Females	54	14	12	10
Statistics	$\chi^2=18$, $p<0.001$			
Age (years)				
Mean	67	56	55	41
Median	69	56	56	39
Range	43–86	18–81	26–78	29–52
25th percentile	60	50	44	35
75th percentile	75	66	71	49
Statistics	Kruskal-Wallis test: $\chi^2=60$, $p<0.001$			
Comorbidity (cases no.)				
Diabetes mellitus	79	30	24	0
RGT	12	3	0	0
Normal	35	28	0	12
Statistics	$\chi^2=46$, $p<0.001$			
CA 19-9 (kU/L)				
Mean	25,133	29	13	14
Median	235	11	11	11
Range	$1–1.3\times10^6$	1–307	0–39	2–32
25th percentile	90	6	6	8
75th percentile	972	30	19	21
Statistics	Kruskal-Wallis test: $\chi^2=86$, $p<0.001$			

RGT, reduced glucose tolerance.

Study 4

This table is taken from Kenny L.C. *et al.* (2005) *Metabolomics* **1**, 227-234. In this study the authors used GC-MS based metabolomics on human serum with the aim of differentiating women with pre-eclampsia[4] from healthy pregnant females.

Demographic data for patients from whom plasma samples were taken

	Normal outcome $n=87$	Preeclampsia $n=87$
Age	30 (19–43)	31 (19–41)
Parity	0 (0–2)	0 (0–2)
BMI (weight/height^2)	25 (19–46)	26 (18–46)
Max (S) BP (mm Hg)	122 (96–147)	162 (138–220)*
Max (D) BP (mm Hg)	80 (60–93)	110 (90–140)*
Delivery gestation (weeks + days)	40 + 4 (34 + 3 to 42 + 0)	37 + 0* (26 + 3 to 41 + 1)
Birth weight (g)	3420 (2380–4420)	2410 (590–4300)*
IBR (centile)	34 (10–99)	8 (0–99)*

Median (range).
Pre-eclampsia vs normal outcome.
*$p<0.0001$.

- *What is the confounder?*
- *How would you test whether this affects the biomarkers discovered?*

[4] pre-eclampsia is pregnancy-induced hypertension

Study 5

Table 1 Clinical parameters for the NCA and TVD patients

	NCA	TVD
Age (years)	57.2 ± 9.0	64.1 ± 7.2
Sex:		
Male (*n*)	7	34
Female (*n*)	23	2
Previous myocardial infarction	1	19
Blood pressure: systolic (mmHg)	141 ± 22	138 ± 23
Diastolic (mmHg)	78 ± 12	75 ± 12
Current smokers (*n*)	2	1
Urea (mM)	5.0 ± 1.2	5.6 ± 1.6
Creatinine (μM)	93 ± 14	108 ± 18
Glucose (mM)	5.2 ± 0.6	5.6 ± 0.9
Total cholesterol (mM)	5.9 ± 1.1	6.2 ± 0.8
HDL-cholesterol (mM)	1.1 ± 0.2	0.8 ± 0.2
LDL-cholesterol (mM)	4.3 ± 1.1	4.5 ± 0.7

^{1}H NMR spectroscopy with PLS was used to predict coronary artery disease in humans as part of trials concerning statins. The table opposite is from Brindle, J.T. *et al*. (2002) *Nature Medicine* **8**, 1439-1445.

The study showed that it was possible to differentiate people with normal coronary arteries (NCA) from patients with triple vessel disease (CVD).

The sample used was human plasma and the key biomarker patterns were in the lipid area of the NMR spectra.

A follow up study (Kirschenlohr, H.L. *et al*. (2006) *Nature Medicine* **12**, 705-710) showed there were two confounders in the original study.

- *Can you suggest what these may be?*

Study 6

A senior consultant has been working on a cure for a rare disease and over the last 20 years has collected 50 valuable serum samples. These could be analysed by FT-IR spectroscopy with the hope of finding a diagnostic pattern. During discussions with your group who wants to work on this important topic there is a suggestion that controls are needed.

- *How would you proceed with this study?*

Study 7

During a drug screen an academic discovers a new magic molecule and this new chemical entity looks like it may be useful for controlling Alzheimer's disease. The group contact a pharmaceutical company to help with Phase I trials[5] and the company is very keen. They set up an initial screen in healthy young men.

- *Do you think this is the correct approach?*

[5] Phase I is typically used to assess how much of the drug is safe to give to someone, what the side effects might be, and how the body copes with the drug (pharmacodynamics).

Study 8

The table below is from Xu, Y. *et al*. (2014) *Metabolomics* **10**, 375–385, where the aim was to use breath analysis (thermal desorption-GC-MS) to differentiate between two lung diseases: COPD[6] *versus* asthma

	COPD	Healthy	Asthma
Age	65.7 (6.7)	55.3 (7.1)	46.1 (14.4)
Gender - % Male	67%	47%	49%
% predicted FEV_1	53.7 (16.3)	102.4 (12.1)	91.5 (22.1)
% predicted FVC	89.7 (17.6)	113.3 (14.6)	106.0 (14.4)
Smokers – current/ex/never	12/27/0	10/0/22	0/0/35
ICS	64%	0%	71%

- *Are the healthy controls here adequate?*
- *What would you suggest that might improve the study outcomes?*

Study 9

A colleague of yours is thinking of using Raman spectroscopy on urine from healthy individuals on 2 different calorie-restricted diets:

(i) Atkins-based Diet which is rich in protein and low in carbohydrates;
(ii) High Carb Diet which is low in protein and high in carbohydrates.

40 healthy individuals have been recruited (50:50 Male:Female).

- *You are asked to help design the study. What would you suggest?*

Study 10

FT-IR or Raman imaging can be used to generate detailed chemical maps from tissues. A colleague of yours wants to use vibrational spectroscopy to investigate tumours in the lung. They ask you about controls for the study.

- *What do you think would be suitable control tissue?*

[6] COPD is chronic obstructive pulmonary disease; a.k.a. emphysema