UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Statistics, data handling and analysis

Fatemeh Torabi

As part of my fellowship of higher education academy Prof. Tom Monie is observing my session today – it is a developmental practice and focused on my approaches and won't include any

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

What you said:

- **Thematic** analysis/Qualitative Coding if possible
- How to go from hypothesis to asking questions in a **survey/questionnaire** to statistical analysis of the answer, with examples
- An example dissertation which has a **survey** in it
- Topic on designing **questionnaires** keeping statistics in mind
- Delphi study related statistics

What am I going to do:

A systematic approach

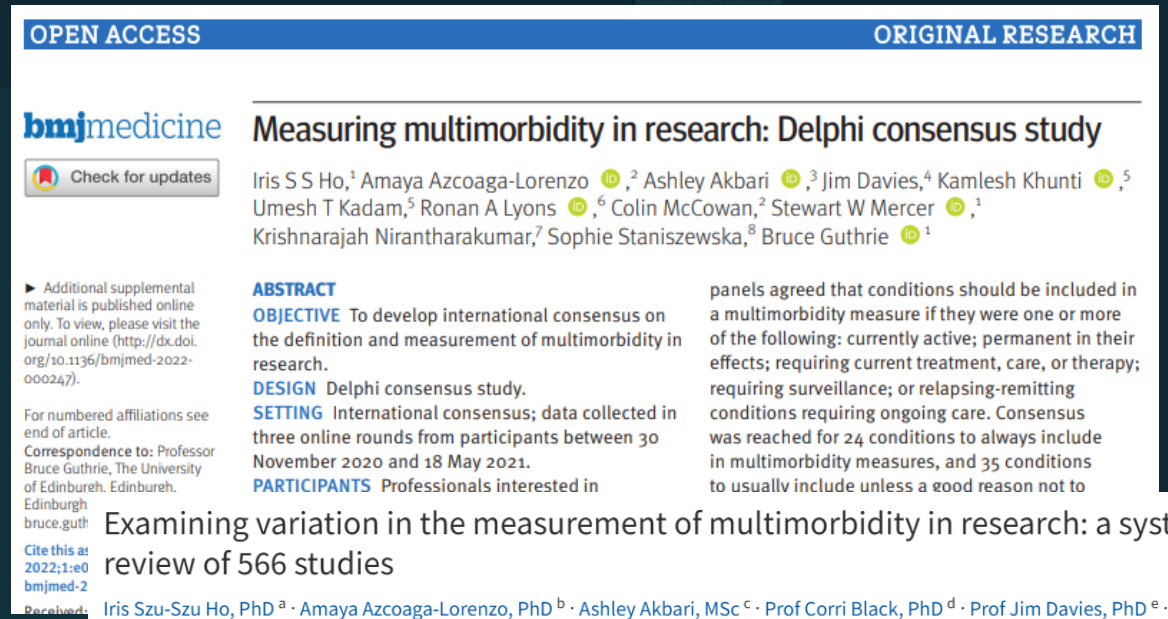
Boost your technical ability

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Delphi method

The Delphi method is a structured communication technique used to gather and consolidate knowledge from a group of experts on a particular topic. It typically involves several rounds of questionnaires or surveys, where experts anonymously provide their opinions, feedback, and predictions. After each round, the responses are summarized and fed back to the participants, who can then revise their answers based on the group's collective insights. This process continues until a consensus or convergence of opinions is reached on the topic being studied.



OPEN ACCESS ORIGINAL RESEARCH

bmjmedicine

Check for updates

Measuring multimorbidity in research: Delphi consensus study

Iris S S Ho,[1] Amaya Azcoaga-Lorenzo,[2] Ashley Akbari,[3] Jim Davies,[4] Kamlesh Khunti,[5] Umesh T Kadam,[5] Ronan A Lyons,[6] Colin McCowan,[2] Stewart W Mercer,[1] Krishnarajah Nirantharakumar,[7] Sophie Staniszewska,[8] Bruce Guthrie[1]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjmed-2022-000247).

For numbered affiliations see end of article.
Correspondence to: Professor Bruce Guthrie, The University of Edinburgh, Edinburgh.
Edinburgh
bruce.guth

Cite this as
2022;1:e0
bmjmed-2
Received

ABSTRACT

OBJECTIVE To develop international consensus on the definition and measurement of multimorbidity in research.

DESIGN Delphi consensus study.

SETTING International consensus; data collected in three online rounds from participants between 30 November 2020 and 18 May 2021.

PARTICIPANTS Professionals interested in

panels agreed that conditions should be included in a multimorbidity measure if they were one or more of the following: currently active; permanent in their effects; requiring current treatment, care, or therapy; requiring surveillance; or relapsing-remitting conditions requiring ongoing care. Consensus was reached for 24 conditions to always include in multimorbidity measures, and 35 conditions to usually include unless a good reason not to

htt

Examining variation in the measurement of multimorbidity in research: a syste review of 566 studies

Iris Szu-Szu Ho, PhD [a] · Amaya Azcoaga-Lorenzo, PhD [b] · Ashley Akbari, MSc [c] · Prof Corri Black, PhD [d] · Prof Jim Davies, PhD [e] · Peter Hodgins, MBBS [a] · Prof Kamlesh Khunti, PhD [f] · Prof Umesh Kadam, PhD [f] · Prof Ronan A Lyons, PhD [c] · Prof Colin McCowa Prof Stewart Mercer, PhD [a] · Krishnarajah Nirantharakumar, MD [g] · Prof Bruce Guthrie, PhD [a] Show less

Affiliations & Notes · Article Info

Download PDF · Cite · Share · Set Alert · Get Rights · Reprints

Show Outline

Summary

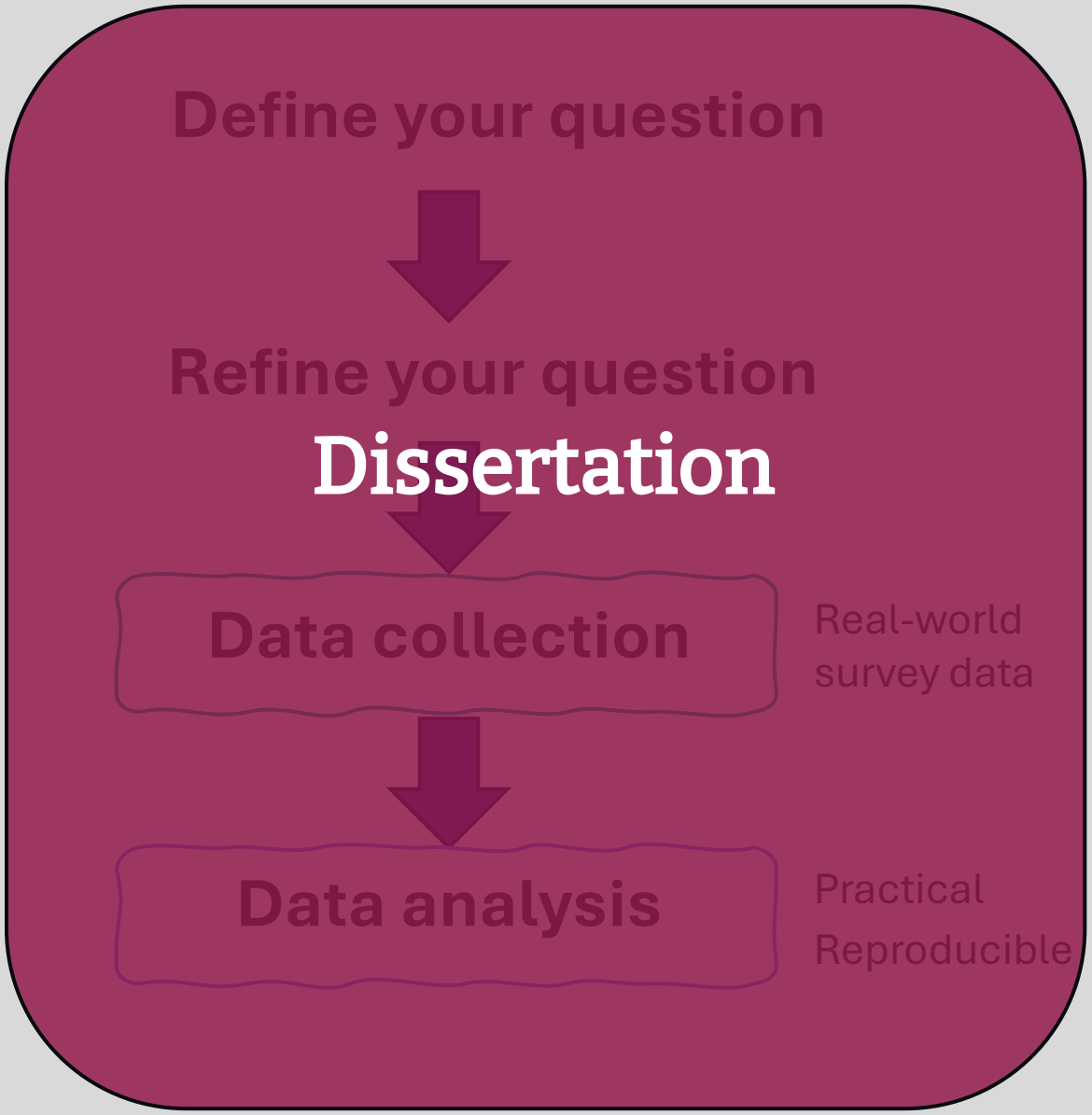
Background

A systematic understanding of how multimorbidity has been constructed and measured is unavailable. This review aime the definition and measurement of multimorbidity in peer-reviewed studies internationally.

https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(21)00107-9/fulltext

Fatemeh.Torabi@ice.cam.ac.uk

Approach



UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Fatemeh.Torabi@ice.cam.ac.uk

Learning outcomes

By the end of this session, you will be able to:

- **Use FINER and PICO frameworks to define a focused, and clinically relevant research question.**
- **Handle real-world survey data (Welsh Health Survery)**
- **Perform exploratory data analysis (EDA) using R, including data wrangling, visualisation, and basic statistical modelling.**
- **Apply reproducible analytical workflows in R using tools like R Markdown.**

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Develop your vision for how these might be applicable in your practice

Fatemeh.Torabi@ice.cam.ac.uk

FINER

Refine

Feasible

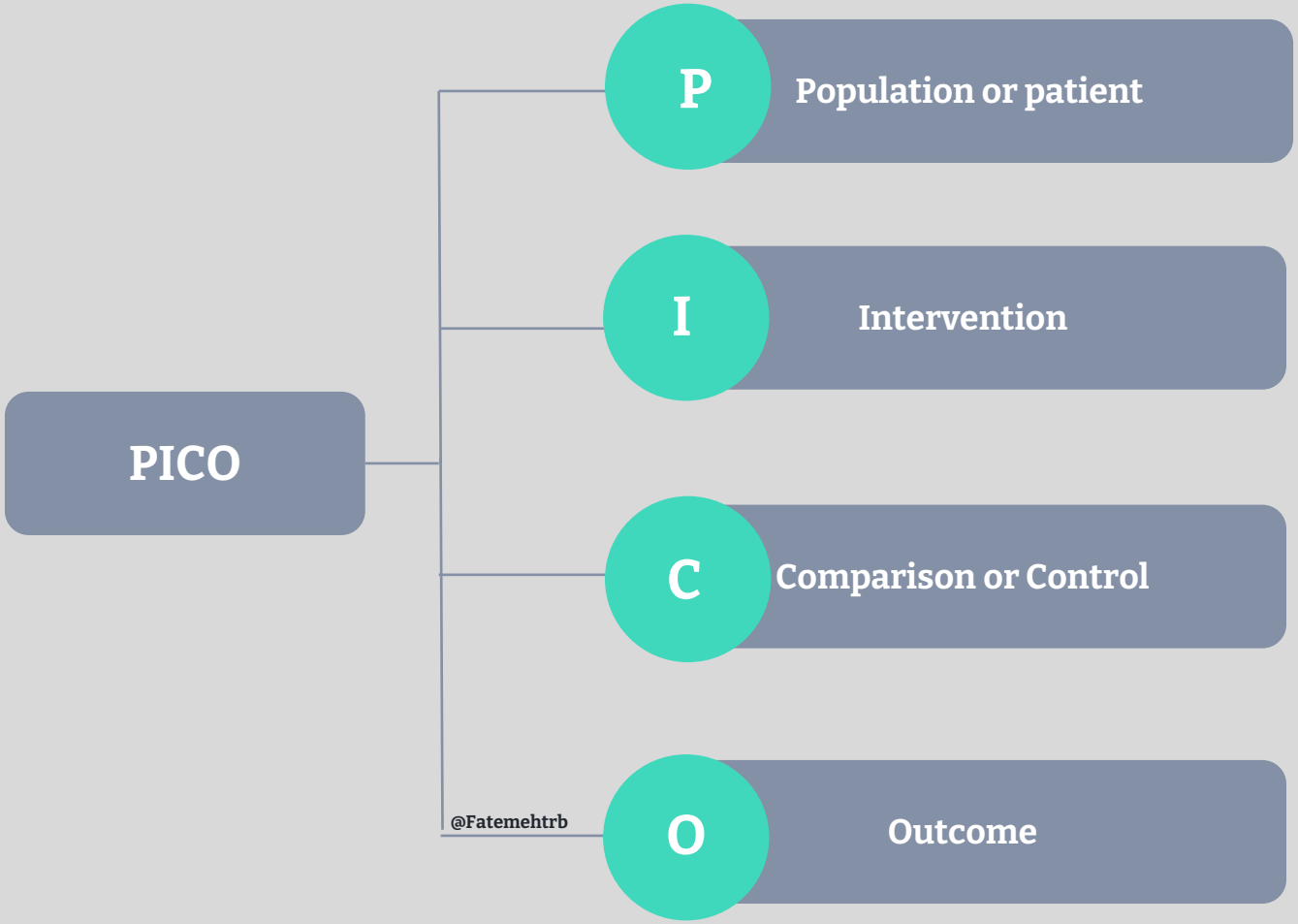
Interesting

Novel

Ethical

Relevant

PICO model



UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Health survey data

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Welsh Health Survey

The survey provides information about the health and health-related lifestyle of people living in Wales. Since 2016 WHS got merged into National Survey for Wales which is a large-scale survey of adults in Wales, covering a range of topics such as wellbeing, people's views on public services and etc.

Self-reported outcomes

Clinician-reported outcomes

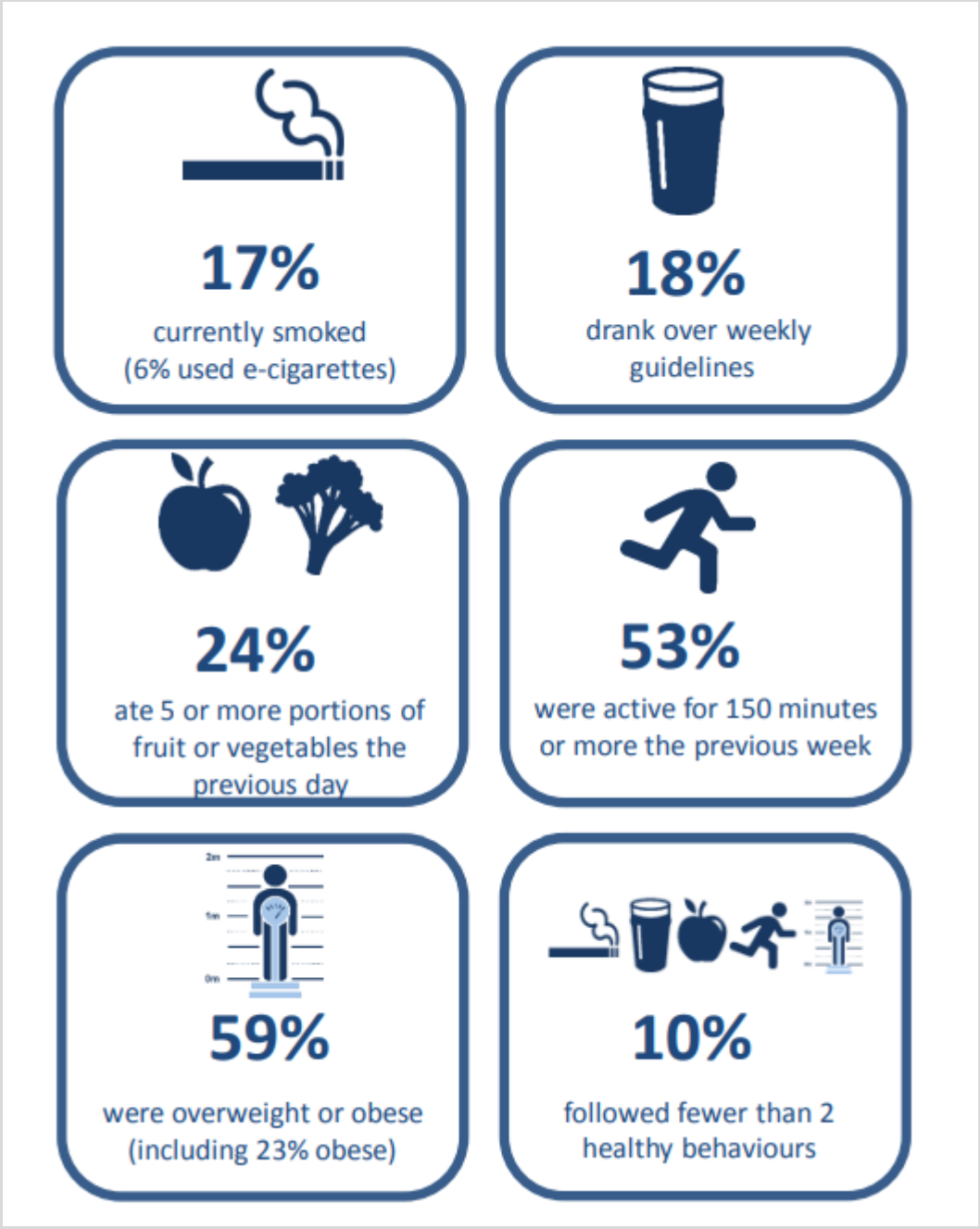
Observer-reported outcomes

Performance-based outcomes

EHR-derived outcomes

Ref: https://statswales.gov.wales/Catalogue/Health-and-Social-Care/Welsh-Health-Survey

Fatemeh.Torabi@ice.cam.ac.uk



UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Fatemeh.Torabi@ice.cam.ac.uk

Does wellbeing in the adult population vary by Body Mass Index (BMI), employment status and general health status?

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Fatemeh.Torabi@ice.cam.ac.uk

D1 | wbAnx

wbSatis	wbWorth	wbHappy	wbAnx	wt_adult	Wellscore	Sex	Age5yrm	health	Work	qualhi	wbsatis1	wbsatis2	w
9	8	8	0	0.790776	20	1	12	3	9	-9	4	3	
8	8	9	1	1.339469	17	2	4	3	2	1	3	2	
0	0	0	10	0.890854	29	2	6	1	1	1	1	1	
5	5	5	5	0.888505	23	1	7	2	2	1	2	1	
5	5	5	8	0.884001	28	2	13	2	9	1	2	1	
6	6	7	0	0.981789	35	1	7	1	2	1	2	1	
7	6	7	7	0.942248	33	2	6	1	11	1	3	2	
5	6	7	0	0.701808	36	1	13	1	9	-9	2	1	
8	9	9	1	0.662964	40	2	12	1	10	0	3	2	
8	8	7	0	0.916824	24	2	13	2	9	-9	3	2	
8	8	10	0	0.935945	22	1	5	1	2	1	3	2	
10	10	10	0	0.925302	33	2	4	1	2	1	4	3	
5	5	6	2	1.030693	-8	1	8	1	2	1		1	
6	7	6	7	1.042827	13	2	6	3	8	0	2	1	
8	8	9	2	1.084329	0	2	1	1	1	1	3	2	
-9	-9	-9	-9	1.165355	36	1	1	1	11	1	-9	-9	
4	4	2	7	1.094714	40	2	8	1	2	0	1	1	
0	0	0	10	1.037683	40	1	9	1	2	1	1	1	
8	8	10	0	1.366535	30	1	3	1	2	1	3	2	
10	10	10	0	0.350814	30	2	9	2	2	1	4	3	
10	10	10	0	0.428951	30	2	13	3	9	1	4	3	
7	7	8	2	0.583863	36	1	3	1	2	2	3	2	
7	8	7	2	0.430569	20	2	8	1	10	0	3	2	
3	7	10	0	0.392529	25	1	10	2	8	0	1	1	
9	10	9	2	0.475727	36	2	3	1	2	1	4	3	
5	5	5	5	0.603534	17	2	6	2	-9	-9	2	1	
5	5	5	0	0.731254	29	2	2	1	2	1	2	1	
10	8	8	0	0.812179	35	1	2	1	2	1	4	3	
0	5	5	3	0.527179	31	1	7	1	2	1	1	1	
6	6	8	1	0.476797	30	2	13	2	10	1	2	1	
9	8	8	0	0.488298	30	2	5	1	6	1	4	3	
7	7	7	0	0.559801	13	1	1	1	2	1	3	2	
6	10	6	2	0.397941	34	2	7	1	2	2	2	1	
7	8	8	3	0.429367	35	1	6	1	2	2	3	2	
3	3	5	8	0.428127	40	2	13	1	9	-9	1	1	
8	9	9	2	0.337782	31	2	10	1	11	1	3	2	
8	9	9	1	0.332907	22	1	11	1	9	1	3	2	
10	10	10	0	0.403798	35	2	13	2	9	1	4	3	
7	8	7	1	0.625615	26	2	2	1	10	0	3	2	
6	7	5	6	0.674796	27	1	2	1	2	0	2	1	
8	9	9	1	0.414451	37	2	5	1	2	-9	3	2	

Sheet1

well 4/4

Wellbeing

Adults only

WBSATIS1 (D) Life satisfaction.
WBWORTH1 (D) Worthwhile.
WBHAPPY1 (D) Happy yesterday.

1 Very low (0-4)
2 Low (5-6)
3 Medium (7-8)
4 High. (9-10)

```
SPSS Syntax
recode wbsatis (0 thru 4=1)(5,6=2)(7,8=3)(9,10=4)(ELSE=COPY) INTO wbsatis1.
recode wbworth (0 thru 4=1)(5,6=2)(7,8=3)(9,10=4)(ELSE=COPY) INTO wbworth1.
recode wbhappy (0 thru 4=1)(5,6=2)(7,8=3)(9,10=4)(ELSE=COPY) INTO wbhappy1.
variable labels wbsatis1 '(D) Life satisfaction'.
variable labels wbworth1 '(D) Worthwhile'.
variable labels wbhappy1 '(D) Happy yesterday'.
value labels wbsatis1 wbworth1 wbhappy1
       1       'Very low (0-4)'
       2       'Low (5-6)'
       3       'Medium (7-8)'
       4       'High (9-10)'.
```

WBSATIS2 (D) Life satisfaction.
WBWORTH2 (D) Worthwhile.
WBHAPPY2 (D) Happy yesterday.

1 Low (0-6)
2 Medium (7-8)
3 High (9-10).

```
recode wbsatis (0 thru 6=1)(7,8=2) (9,10=3)(ELSE=COPY) INTO wbsatis2.
recode wbworth (0 thru 6=1)(7,8=2) (9,10=3)(ELSE=COPY) INTO wbworth2.
recode wbhappy (0 thru 6=1)(7,8=2) (9,10=3)(ELSE=COPY) INTO wbhappy2.
variable labels wbsatis2 '(D) Life satisfaction'.
variable labels wbworth2 '(D) Worthwhile'.
variable labels wbhappy2 '(D) Happy yesterday'.
value labels wbsatis2 wbworth2 wbhappy2
       1       'Low (0-4)
       2       'Medium (5-10)'
       3       'High (9-10)'.
```

WBSATIS3 (D) Life satisfaction.
WBWORTH3 (D) Worthwhile.
WBHAPPY3 (D) Happy yesterday.

1 Very low (0-4)
2 Not very low (5-10).

```
recode wbsatis (0 thru 4=1)(5 thru 10=2)(ELSE=COPY) INTO wbsatis3.
recode wbworth (0 thru 4=1)(5 thru 10=2)(ELSE=COPY) INTO wbworth3.
recode wbhappy (0 thru 4=1)(5 thru 10=2)(ELSE=COPY) INTO wbhappy3.
variable labels wbsatis3 '(D) Life satisfaction'.
variable labels wbworth3 '(D) Worthwhile'.
variable labels wbhappy3 '(D) Happy yesterday'.
value labels wbsatis3 wbworth3 wbhappy3
       1       'Very low (0-4)'
       2       'Not very low (5-10)'.
```

Does wellbeing in the adult population vary by Body Mass Index (BMI), employment status and general health status?

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Fatemeh.Torabi@ice.cam.ac.uk

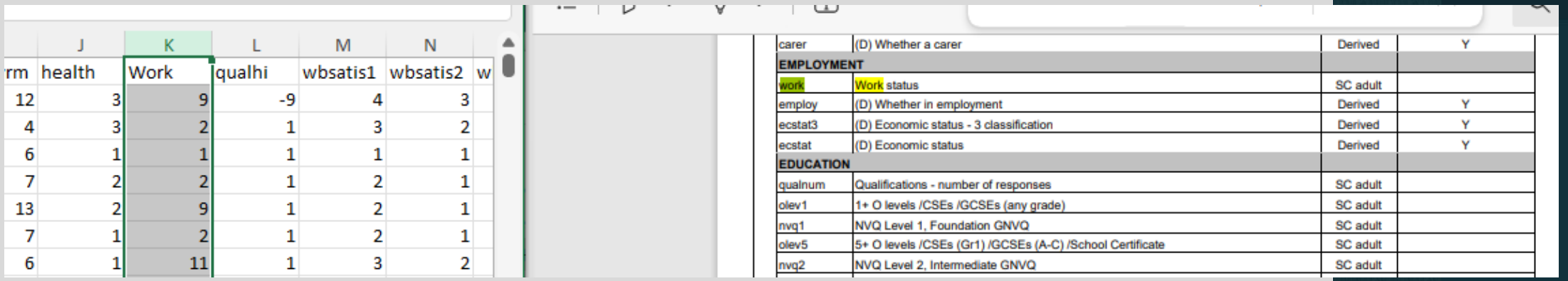
General health status

AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	
th	genhlth	healthy	worshlth	exhlth	llti2	smokee	smokex	nodrink	alcagbi	a
0	3	3	3	2	3	1	1	-1	1	
1	4	5	1	5	2	1	1	-1	0	
1	5	5	1	5	1	1	1	-1	0	
1	5	5	1	5	1	1	1	-1	0	
1	4	3	3	5	2	1	0	-1	0	
0	3	2	3	5	3	0	0	-1	1	
0	3	2	3	2	3	0	0	-1	1	
0	3	2	4	2	1	1	0	-1	1	
0	1	3	3	2	3	0	0	-1	0	
0	1	2	5	1	3	0	0	-1	1	
0	2	5	3	5	3	0	0	-1	1	
0	2	1	3	2	3	0	0	-1	1	
1	4	3	2	4	2	0	0	1	0	
0	3	2	3	4	3	1	1	-1	-9	
0	2	2	4	2	3	0	0	-1	0	

Variable	Description
difdocu	It is difficult to see a doctor/ It would have taken too long
livewitu	I have learnt to live with my symptoms/ I put up with my symptoms
seenothu	I have seen someone else about these
seendocu	I have seen a doctor about these symptoms more than 12 months ago
otheru	Other reasons for no treatment
HEALTH AND WELL BEING	
fphth	(D) Fair or poor health - binary
genhlth	SF Self reported health
comphlth	SF Health compared with year ago
vigact	SF Health limits vigorous activities
modact	SF Health limits moderate activities
liftgroc	SF Health limits lifting and carrying groceries
climbsev	SF Health limits climbing several flights of stairs
climbone	SF Health limits climbing one flight of stairs
bend	SF Health limits bending, kneeling or stooping
walkmile	SF Health limits walking more than a mile
walksvyd	SF Health limits walking several hundred yards
walkhdyd	SF Health limits walking one hundred yards
bath	SF Health limits bathing or dressing yourself
physcut	SF Due to physical health cut down time on work or activities in past 4 weeks
physless	SF Due to physical health accomplished less than would like in past 4 weeks
physlim	SF Due to physical health limited in kind of work or activities in past 4 weeks

Fatemeh.Torabi@ice.cam.ac.uk

Employment status



	J	K	L	M	N	
rm	health	Work	qualhi	wbsatis1	wbsatis2	w
12	3	9	-9	4	3	
4	3	2	1	3	2	
6	1	1	1	1	1	
7	2	2	1	2	1	
13	2	9	1	2	1	
7	1	2	1	2	1	
6	1	11	1	3	2	

carer	(D) Whether a carer	Derived	Y
EMPLOYMENT			
work	Work status	SC adult	
employ	(D) Whether in employment	Derived	Y
ecstat3	(D) Economic status - 3 classification	Derived	Y
ecstat	(D) Economic status	Derived	Y
EDUCATION			
qualnum	Qualifications - number of responses	SC adult	
olev1	1+ O levels /CSEs /GCSEs (any grade)	SC adult	
nvq1	NVQ Level 1, Foundation GNVQ	SC adult	
olev5	5+ O levels /CSEs (Gr1) /GCSEs (A-C) /School Certificate	SC adult	
nvq2	NVQ Level 2, Intermediate GNVQ	SC adult	

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Fatemeh.Torabi@ice.cam.ac.uk

Body Mass Index (BMI),

	AA	AB	AC	AD	AE	AF	AG
2	wbanx3	wbanx4	htcm	wtkg	anx	anxbi	fphth
1	1	1	177.8	93	2	0	0
1	1	1	165.1	64.41	2	0	1
3	2	2	149.86	82.56	1	1	1
3	1	2	170.18	72	2	0	1
3	2	2	152.4	69.85	1	1	1
1	1	1	152.4	69.85	2	0	0
3	2	2	157.48	66	-9	-9	0
1	1	1	165.1	63.5	2	0	0
1	1	1	165.1	76.2	2	0	0
1	1	1	160.02	82.56	2	0	0
1	1	1	175.26	80.74	2	0	0
1	1	1	167.64	71.67	2	0	0
2	1	1	157.48	60.33	-9	-9	1
3	2	2	175.26	98	2	0	0
2	1	1	165	59	2	0	0
-9	-9	-9	-9	-9	-9	-9	0

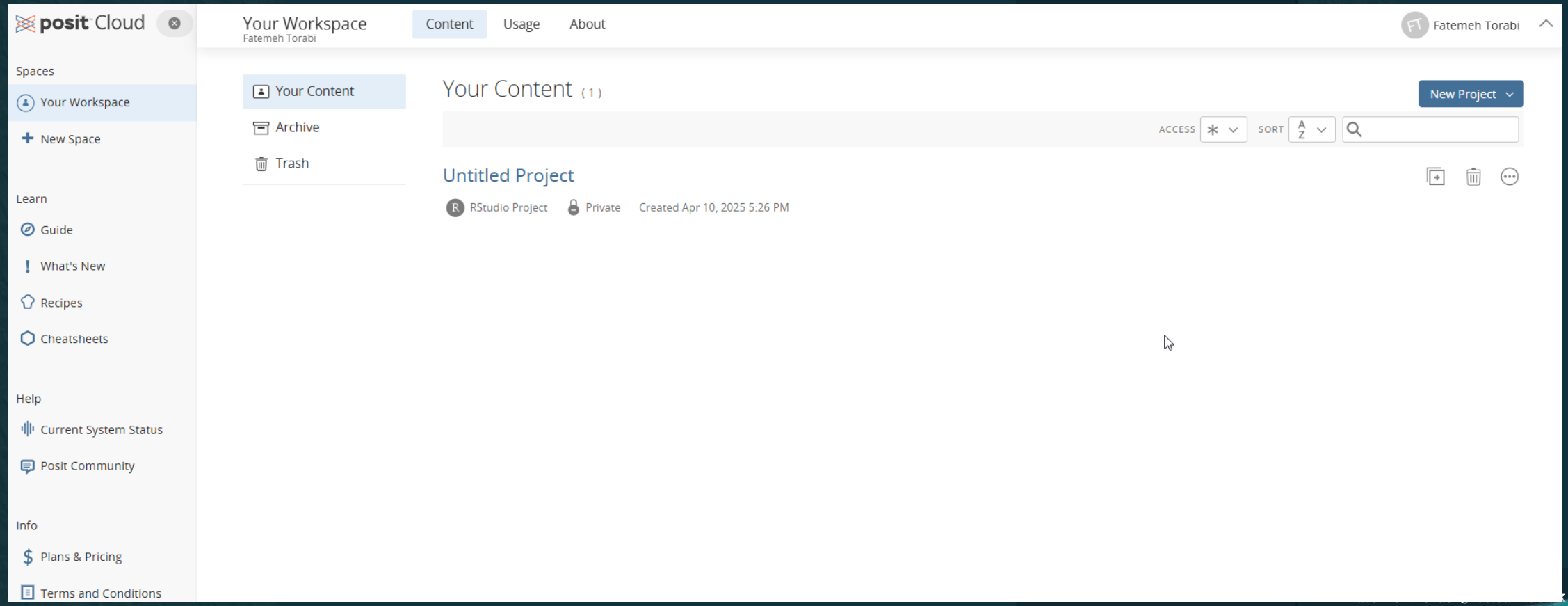

```
data$BMI <- data$wtkg / ((data$htcm/100)^2)
```


UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

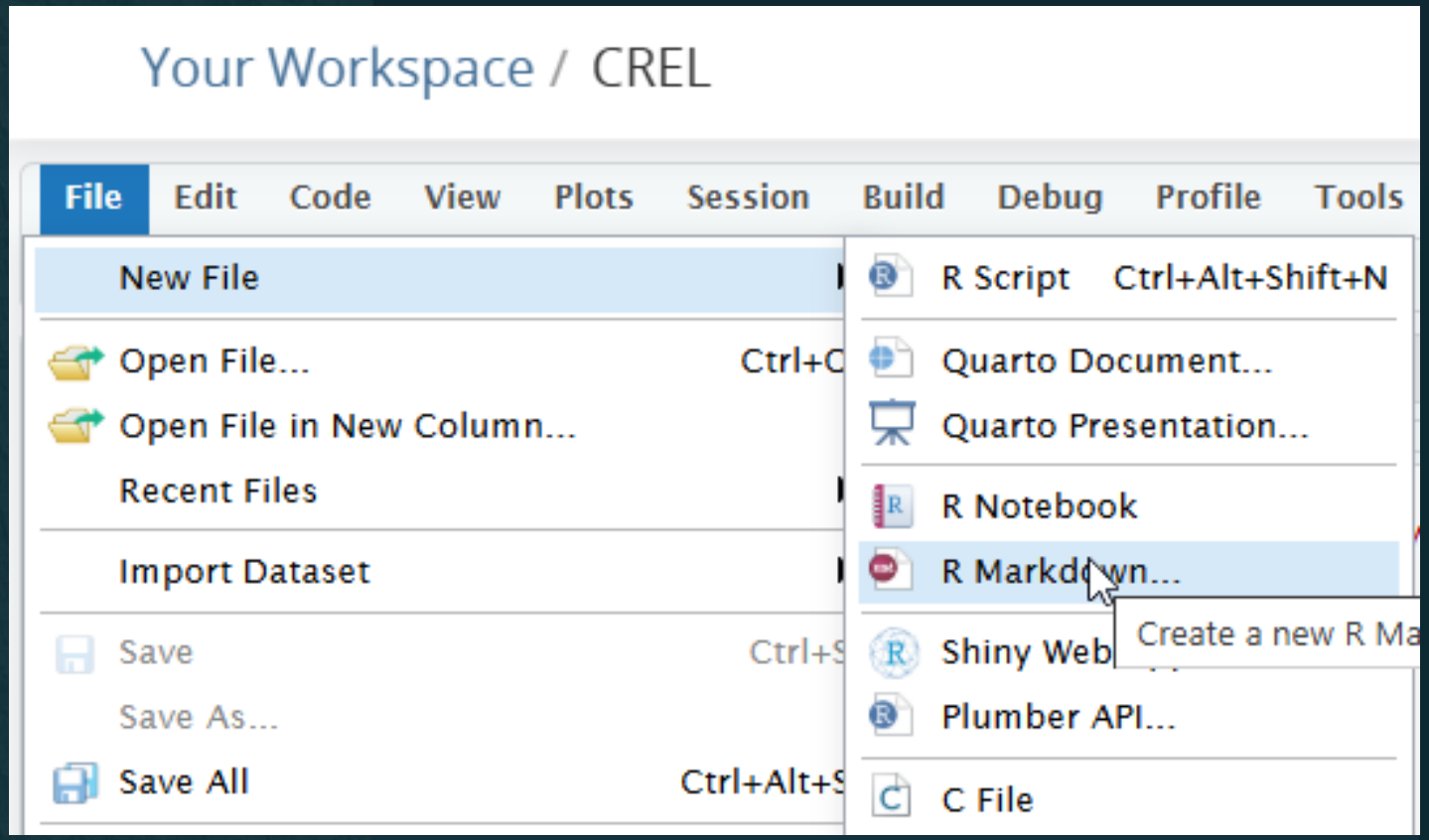
Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

https://posit.cloud/



UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION



Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

When running the EDA report from R cloud server you need to load the data from the server

````
28
29 ```{R, message=FALSE, echo=FALSE}
30 #loadind dataset
31 library(readxl)
32 WHS <- read_excel("CREL Survey sample/WHS.xlsx")
33
34 ```
````

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Download R and Rstudio

https://github.com/FatemehTorabi/EDA-in-R

SCAN ME

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

More resources

- https://cambiotraining.github.io/stats-glm/
- https://rpubs.com/mestrad/1033097

Fatemeh.Torabi@ice.cam.ac.uk

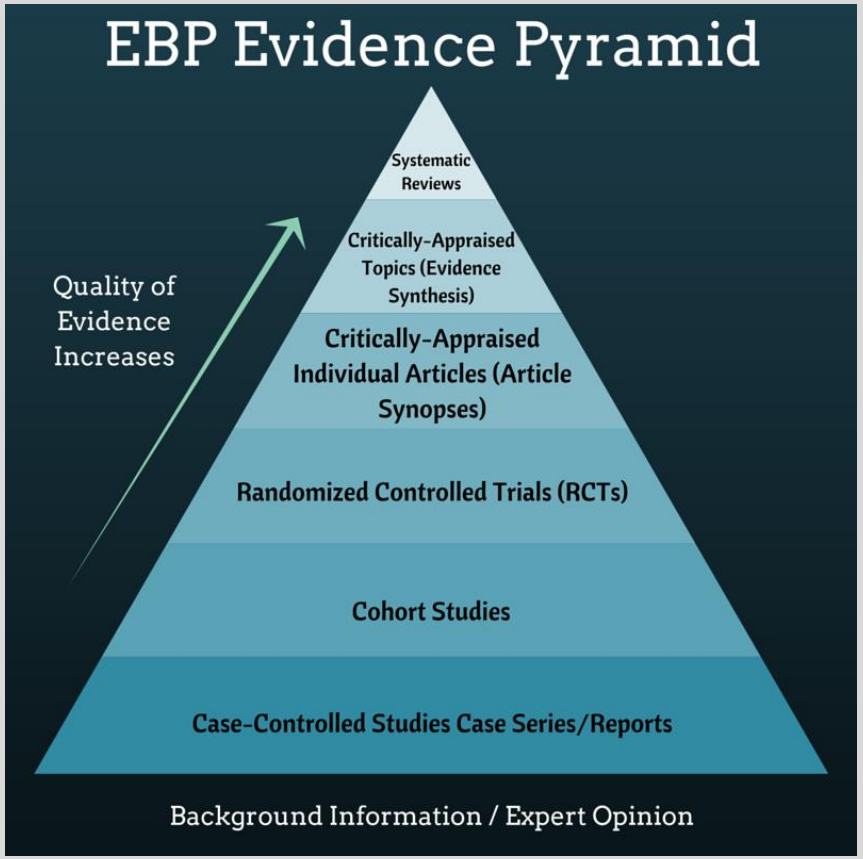
UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Dissertation

Fatemeh.Torabi@ice.cam.ac.uk

Dissertation - key characteristics

- Structured
- Cohesive piece
- Authentic
- Simple
- Complete
- Clear academic writing



UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

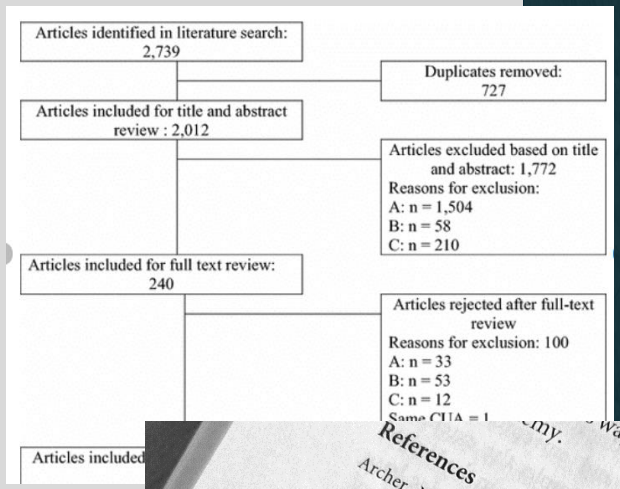
Fatemeh.Torabi@ice.cam.ac.uk

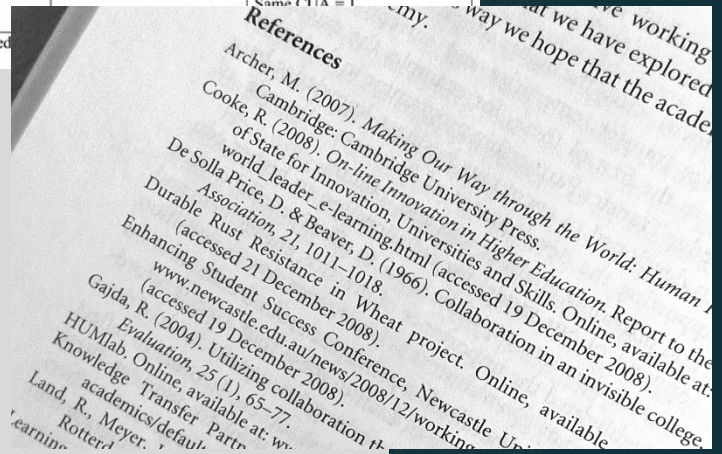
UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

The aim is to provides a strong body of evidence (old and current) and rationale to support your research question

have a well-documented reproducible approach in searching through evidence based.

- Keep record at each stage of your search
- Ensure your screening strategy is clear





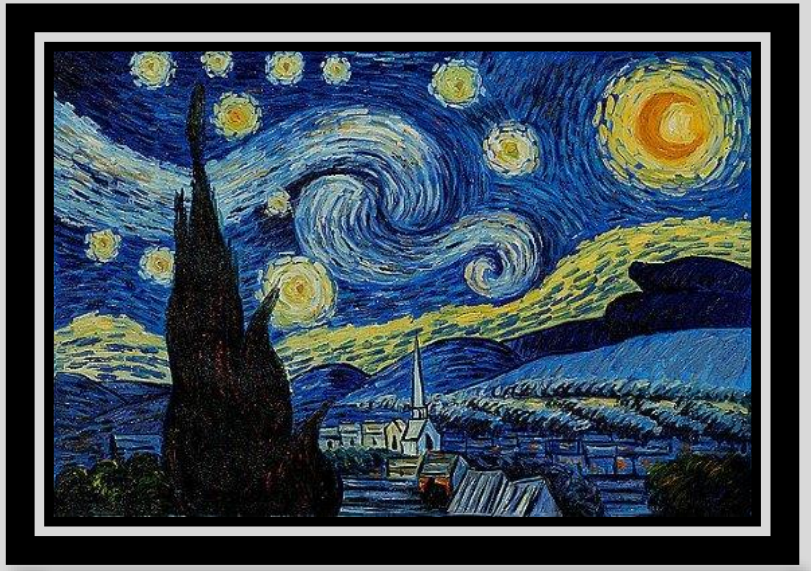
Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Your art piece…

Within defined limits

Fatemeh.Torabi@ice.cam.ac.uk



UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

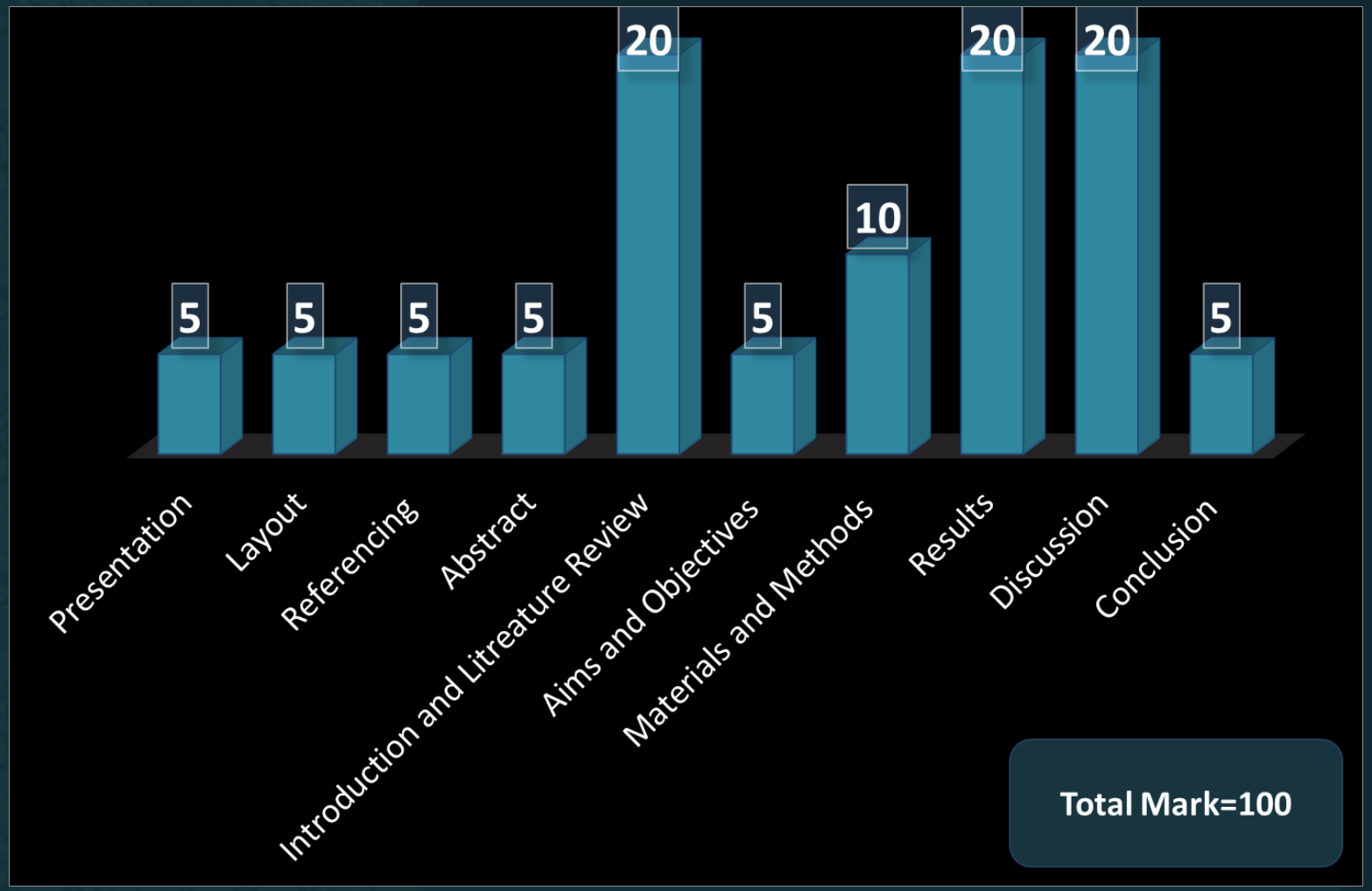
Word limit : 12,000

Defined structure

Data limitations

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION



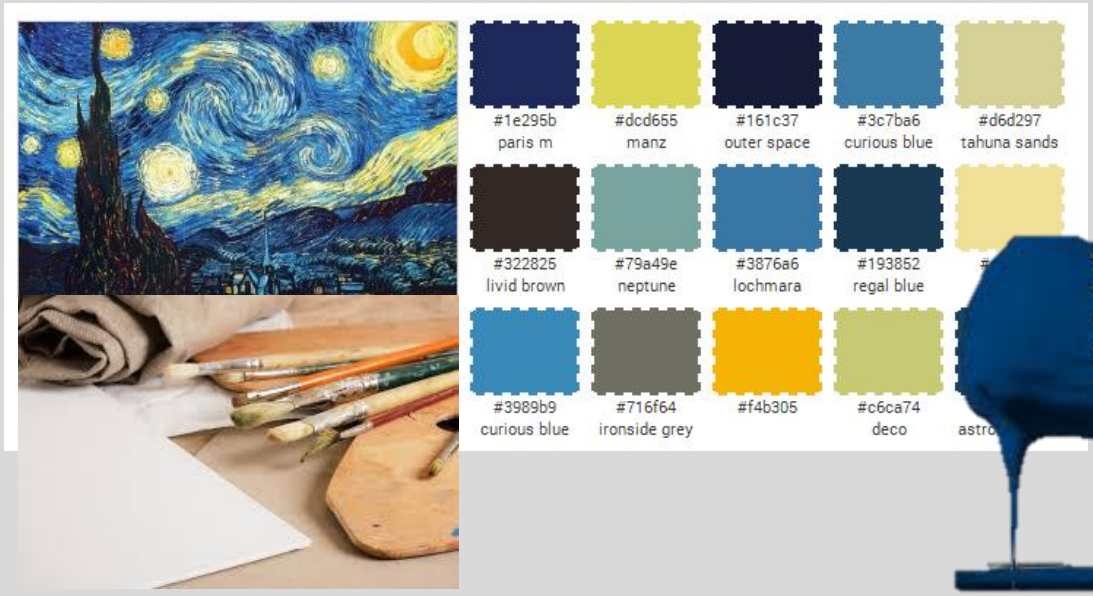
Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Introduction and Literature review (20%)

Fatemeh.Torabi@ice.cam.ac.uk

What is the context?
Where is the gap/need?
How am I addressing this?



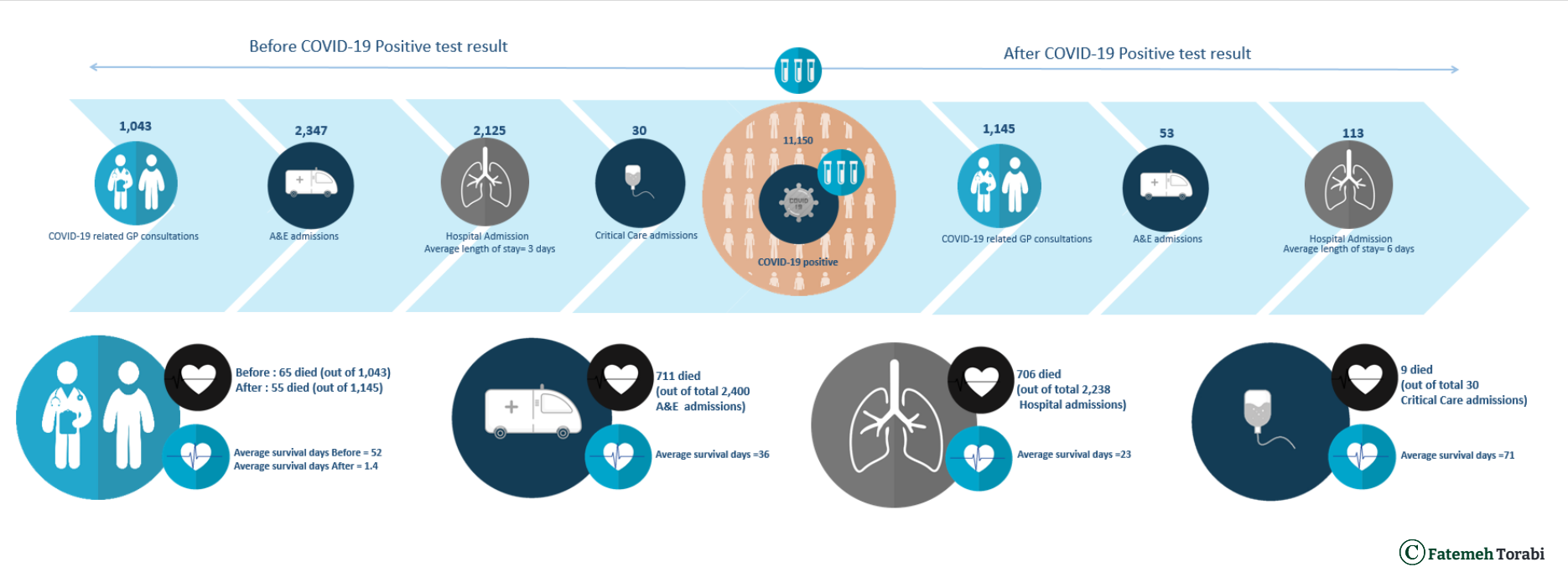
1. Should account for around one third of the word count. (~4000 words)
2. Detailed background with appropriate contemporary and well sourced citations.
3. Balanced and erudite analysis showing a comprehensive level knowledge of the field.
4. Thoughtful approach to gaps in the literature.
5. Relevance to wider health data science landscape.
6. Provides a clear rationale for the research project undertaken.

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Effective use of data to solve problems:



Quality data needs to reach to the right people in a right time

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Materials and Methods (10% Marks)

Fatemeh.Torabi@ice.cam.ac.uk

This will differ depending on the chosen methodology however should include:

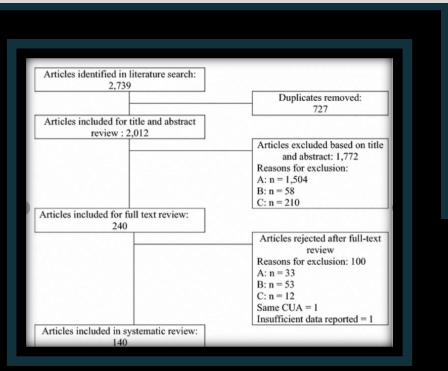
- **Definitions**, where required.
- Data type, sources, composition, specification, and **population** described.
- **Ethical** and legislative aspects.
- Informatics methodology, software and tools used.
- Statistical tools and techniques.
- Approach to visualisation. Tools used, strengths and weaknesses.
- Students should demonstrate **comfort** with **data types and structures**, approaches to informatics, understanding of the **process and limitations for data sharing**, and effective communication and change management.

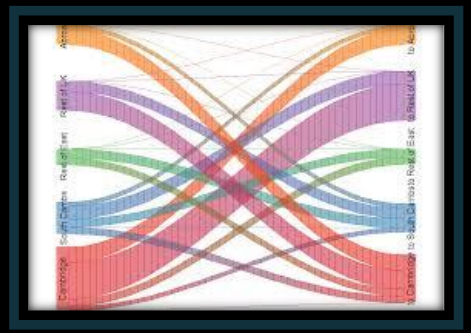
UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Results (20% Marks)

Results are reported systematically and coherently with supporting tables and figures and accurate statistical findings (where appropriate).

Identification of limitations, bias, and confounding to the data and its interpretation.





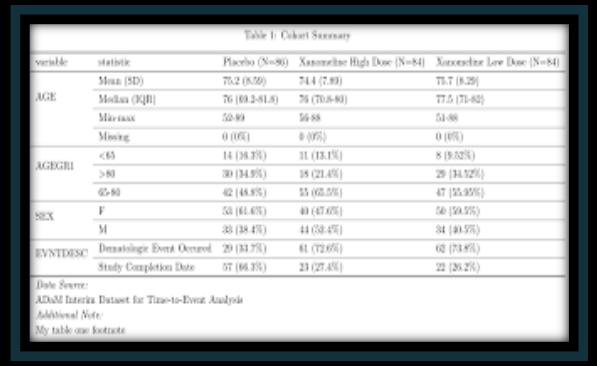
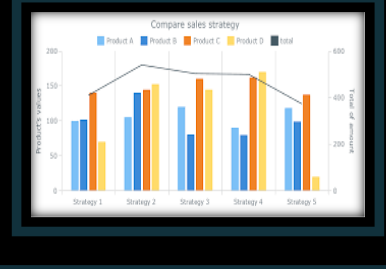
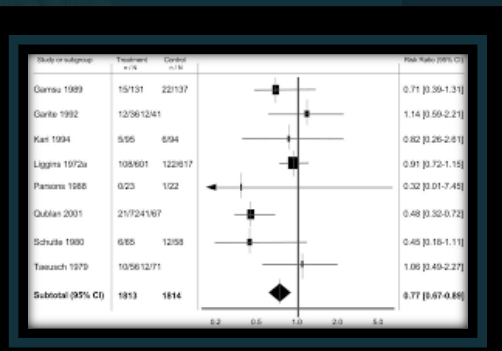


Table 1: Cohort Summary

variable	statistic	Placebo (N=86)	Xanomeline High Dose (N=84)	Xanomeline Low Dose (N=84)
AGE	Mean (SD)	75.2 (8.59)	74.4 (7.89)	75.7 (8.29)
	Median (IQR)	76 (69.2-81.8)	76 (70.8-80)	77.5 (71-82)
	Min-max	52-89	56-88	51-88
	Missing	0 (0%)	0 (0%)	0 (0%)
AGEGR1	<65	14 (16.3%)	11 (13.1%)	8 (9.52%)
	>80	30 (34.9%)	18 (21.4%)	29 (34.52%)
	65-80	42 (48.8%)	55 (65.5%)	47 (55.95%)
SEX	F	53 (61.6%)	40 (47.6%)	50 (59.5%)
	M	33 (38.4%)	44 (52.4%)	34 (40.5%)
EVNTDESC	Dermatologic Event Occurred	29 (33.7%)	61 (72.6%)	62 (73.8%)
	Study Completion Date	57 (66.3%)	23 (27.4%)	22 (26.2%)

Data Source:
ADaM Interim Dataset for Time-to-Event Analysis
Additional Note:
My table one footnote





UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

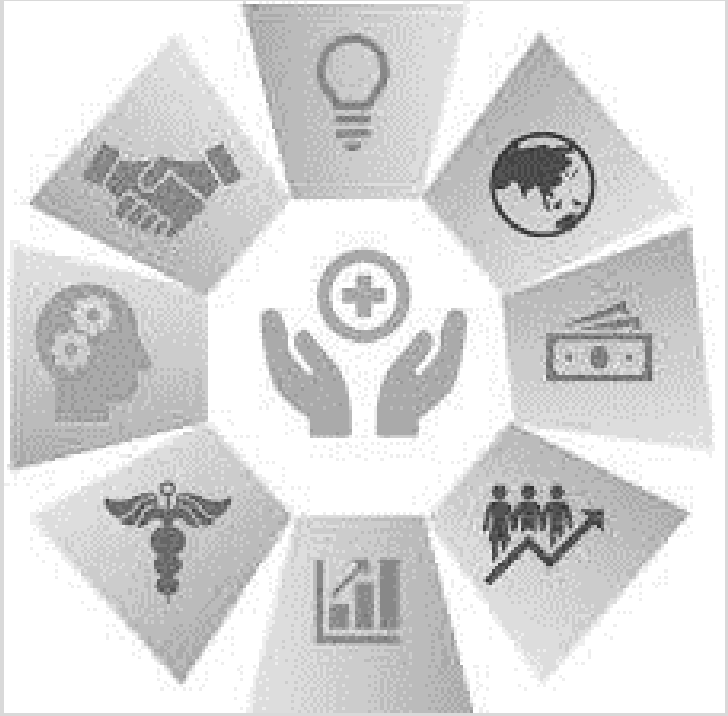
Discussion (20% Marks)

- Critical discussion of the findings with clear evidence of an understanding of how the project sits within the wider landscape of healthcare data.
- The answer to the “So What?” question
- Strengths – limitations
- Future work

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Conclusion (5% Marks)



- Two or three key findings of the study.
- Its impact in policy, practice etc.

Fatemeh.Torabi@ice.cam.ac.uk

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CONTINUING EDUCATION

Thank you!

Fatemeh.Torabi@ice.cam.ac.uk