Applied Statistics Lab Project 1: Epidemiology and Lifestyle Determinants of High Blood Pressure in the United States

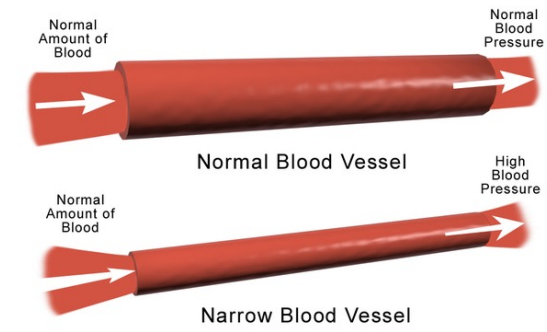Introduction to Statistical Thinking and Data Analysis
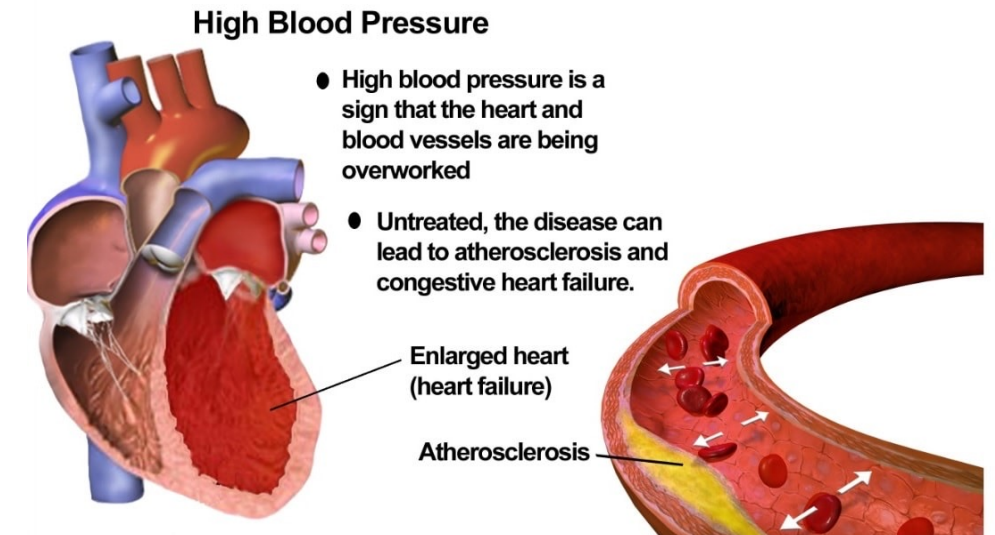MSc in Epidemiology / Health Data Analytics
Autumn 2022

17 October 2022

Project 1 – Topic

Epidemiology and Lifestyle Determinants of High Blood Pressure in the United States

BLOOD PRESSURE CATEGORY	SYSTOLIC mm Hg (upper number)		DIASTOLIC mm Hg (lower number)
NORMAL	LESS THAN 120	and	LESS THAN 80
ELEVATED	120 – 129	and	LESS THAN 80
HIGH BLOOD PRESSURE (HYPERTENSION) STAGE 1	130 – 139	or	80 – 89
HIGH BLOOD PRESSURE (HYPERTENSION) STAGE 2	140 OR HIGHER	or	90 OR HIGHER
HYPERTENSIVE CRISIS (consult your doctor immediately)	HIGHER THAN 180	and/or	HIGHER THAN 120

Normal Amount of Blood

Normal Blood Pressure

Normal Blood Vessel

High Blood Pressure

Normal Amount of Blood

Narrow Blood Vessel

Blood Pressure Blood Flow

High Blood Pressure

- High blood pressure is a sign that the heart and blood vessels are being overworked
- Untreated, the disease can lead to atherosclerosis and congestive heart failure.
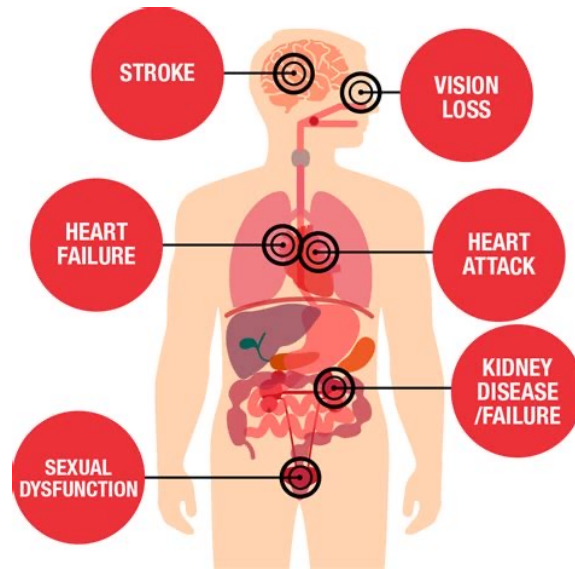
Enlarged heart (heart failure)

Atherosclerosis

Project 1 – Topic

Epidemiology and Lifestyle Determinants of High Blood Pressure in the United States

Serious health risk:
1. Stroke, CVD
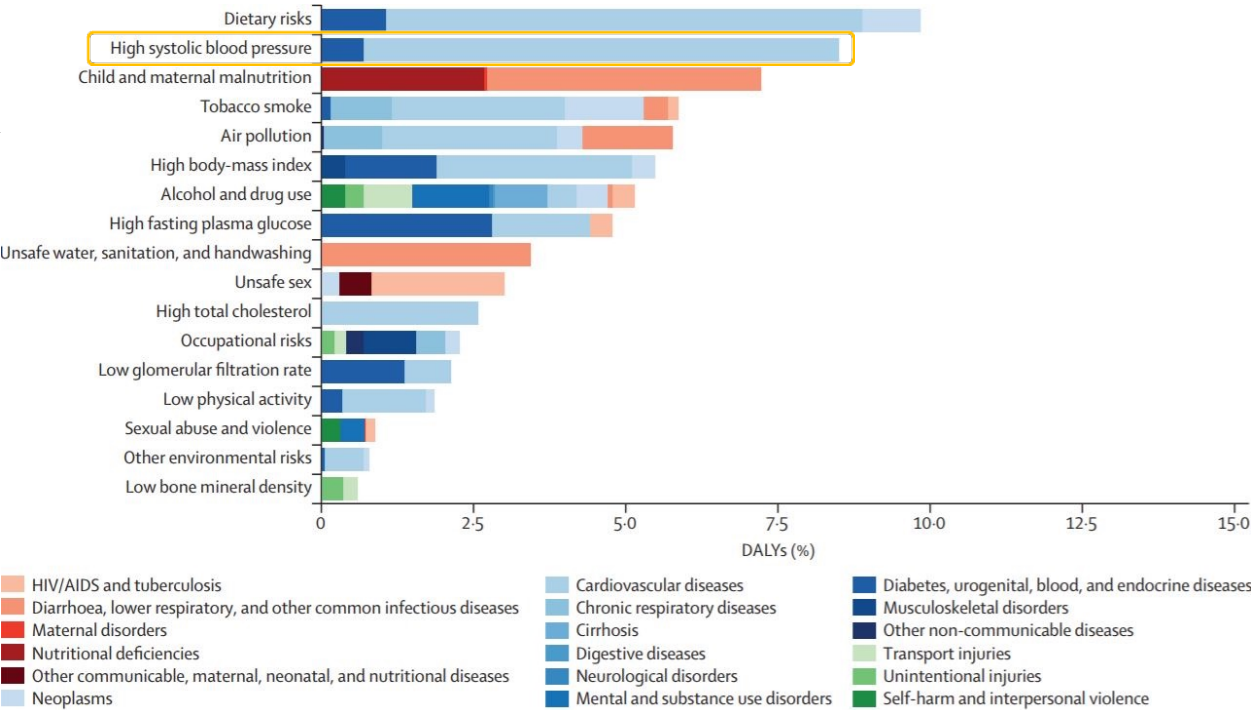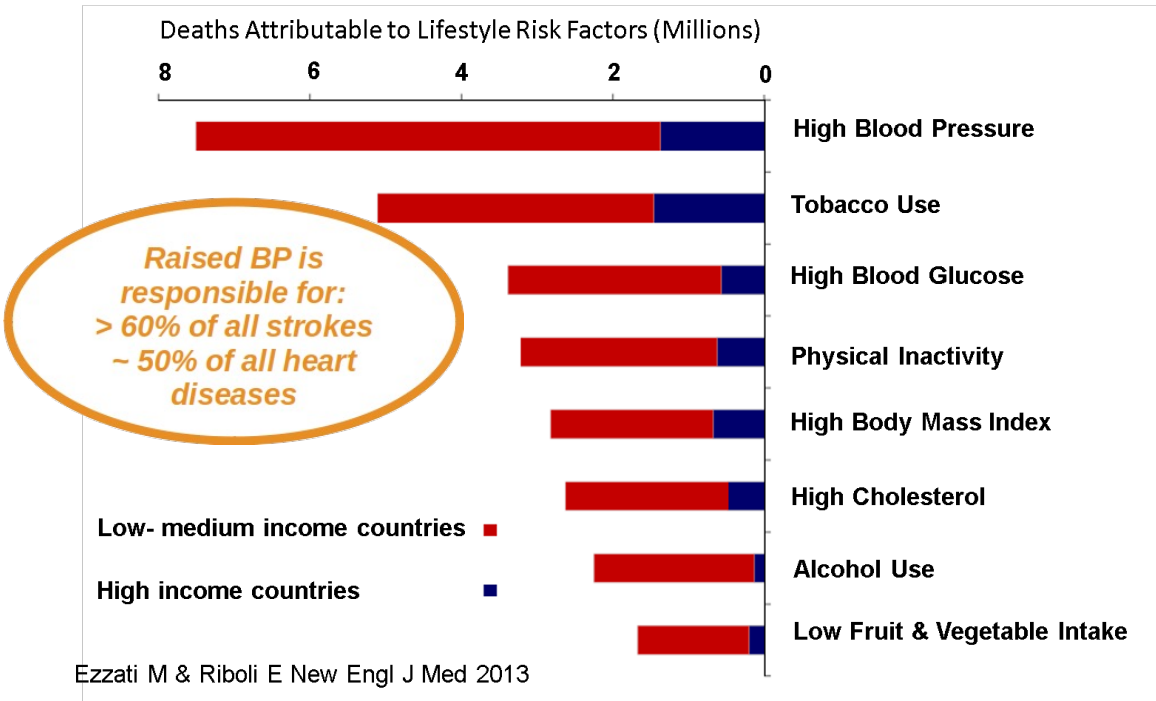2. Kidney disease
3. Vision loss
4. Vascular dementia

1. = often fatal or very debilitating

- Lead risk factor of preventable morbidity/ mortality worldwide
- Treatment
  - Medication: ACE inhibitors, ARBs, diuretics,…
  - Life-style change: weight loss, health diet, quit smoking...

Project 1 – Topic

Epidemiology and Lifestyle Determinants of High Blood Pressure in the United States

Deaths Attributable to Lifestyle Risk Factors (Millions)

Raised BP is responsible for: > 60% of all strokes ~ 50% of all heart diseases

Ezzati M & Riboli E New Engl J Med 2013

Forouzanfar et al, Lancet 2015

Project 1 – Topic

Epidemiology and Lifestyle Determinants of High Blood Pressure in the United States

- Lead risk factor of preventable morbidity/ mortality worldwide
- Treatment
  - Medication: ACE inhibitors, ARBs, diuretics,…
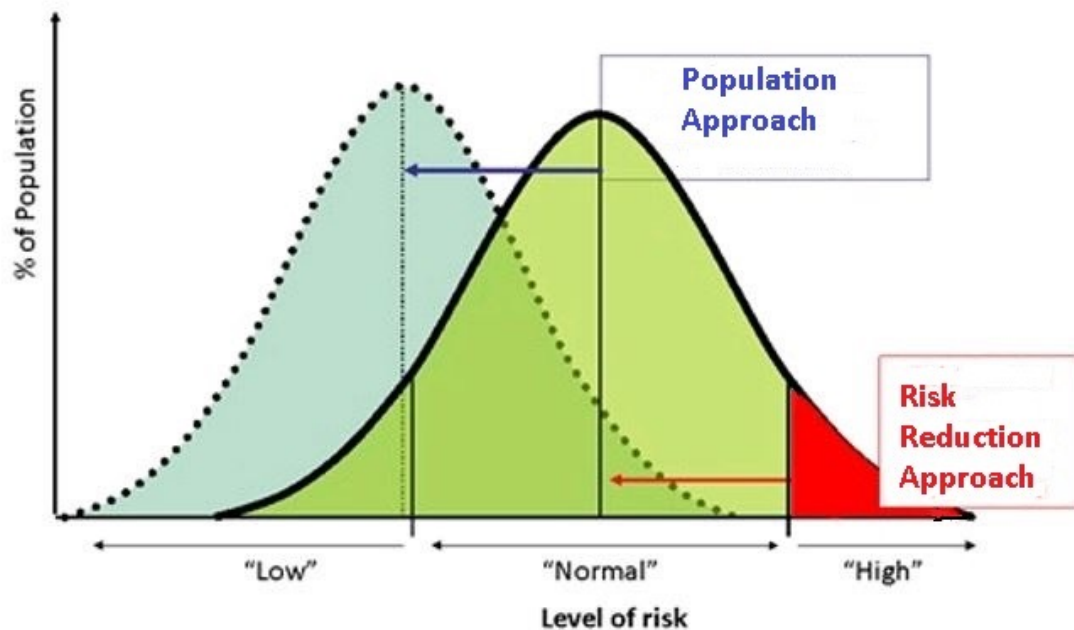 - **Life-style change: weight loss, health diet, quit smoking...**

Project 1 – Aims

Epidemiology and Lifestyle Determinants of High Blood Pressure in the United States

1. **Describe Epidemiology of blood pressure (BP)** in Adults in USA dataset
   1. Mean BP estimate
   2. Identify demographic groups with higher BP average

2. **Identify lifestyle risk factors**
   1. Targetable by public health campaigns,
   2. Aim: reduce prevalence of hypertension

Project 1 – Data set

Epidemiology and Lifestyle Determinants of High Blood Pressure in the United States

- Data set: National Health and Nutrition Examination Survey (NHANES)
  - 75 variables
  - 10,000 subjects
  - Collected between 2009 - 2012

Access: in Rstudio

```
install.package("NHANES")
 library(NHANES)
 data(NHANES)
```

nhanes

National Health and Nutrition Examination Survey

Project 1 – Timeline

Week 1: 17 October (today) - Develop analysis plan
- Specify research questions
  - Identify 2-3 specific hypotheses
- Define dataset
  - Inclusion or exclusion criteria
- Identify variables to use
  - Exploratory data analysis
  - Quality of measurement and data missingness
- Agree on statistical methods
- Define table shells for reporting results

After 45 minutes:
- Each group present 1 slide stating your hypotheses.

15 minutes before session end:
- Each group present 1 slide defining outcome and exposure variables for each hypothesis.

Project 1 – Timeline

Week 2: 24 October - Statistical analysis + interpretation
- Carry out your statistical analysis plan developed in Week 1.
- Interpret results with respect to research questions and hypotheses.
- Check any important assumptions
  - Sensitivity analyses
  - Enumerate any important limitations
- Contextualize your findings compared to other evidence.
- Identify key recommendations

For 3rd and final session:
- Prepare a 10-minute presentation of your results, including:
  - Background, research question, and hypotheses.
  - Methods: variables used and statistical analyses.
  - Results
  - Interpretation and conclusions

Project 1 – Timeline

Week 3: 31 October - Present results

- Present your work in groups
  - One person 'share screen' to present slides.
  - **Each group member** expected to participate in presentation.
- Debrief project 1
  - One thing your group did well.
  - One thing you would do differently next time.
  - One thing you liked about a different group's presentation

Structure of presentation day:

- 10 minutes at the start of the session for any final coordination of your presentation.
- 10 minutes presentation for each group
  - followed by 3 minutes for questions + discussion from peers.
- 10 minutes debrief in small groups
- 20 minutes sharing debriefs with all groups

Imperial College London

Criteria	Marks (Total 40)
Introduction	
Research question and population of interest identified.	3
Clear motivation for why research is important and rationale for study.	3
Specific objectives and testable hypotheses articulated.	3
Methods	
Dataset, inclusion criteria, and data processing clearly described and suitable to address question.	3
Clear description of analysis plan to adjudicate hypotheses, including clear identification of primary outcome and exposure variables.	3
Clear description of choice of statistical model to address the hypothesis and why chosen.	3

Criteria	Marks
Results	
Appropriate presentation of descrivtive statistics; communicated understanding of key relevant features of the study population.	3
Presentation of results inferential statistical analyses proves clear and concise answers to study hypotheses	3
Clear and focused figures and tables used effectively to aid communication of results narrative.	3
Accurate and elegant presentation and interpretation of statistical outcomes and sensitivity analyses.	3
Conclusions	
Clear statement and interpretation of results linked to overall study aims and hypotheses	3
Findings are contextualized and implications for public health practice and research articulated	3
Presentation in allotted time and well paced	2
Overall quality and clarity of presentation slides	2

Steps in a statistical analysis

1. Identify the research question and hypothesis

2. Define target population

3. Select the dataset and define inclusion/exclusion criteria.

4. Identify variables to be used in main analysis (outcome, main exposure, stratifying variables)

5. Define key shell tables

Steps in a statistical analysis

6. Conduct exploratory data analysis: understand your dataset.

7. Conduct inferential statistics: answer your research question

8. Check assumptions and sensitivity analysis: are the findings robust

9. Interpret the results: state the conclusion to the research question
   – Should be as close to a 'yes'/'no' answer to your hypothesis as possible.

10. Contextualize the findings: how do the findings compare to existing evidence.

11. Implications for policy and practice

Communicating statistical analyses

- Communication of statistical findings almost *never* involves computer code or screen output from statistical programs.

- The exact steps and processes of statistical analysis (R commands run, assumptions checked, etc.) are usually not the focus reporting on an analysis.

- But it is important that the reporting of results communicates that statistical steps have been done properly.
  - This will also be the standard for assessment in this course.

Helpful notes

- Blood pressure is measured in *'millimeters of mercury'* (mmHg).

- **Systolic blood pressure** is most closely monitored.
  - Most predictive of increased risk for stroke and heart disease.
  - See variable `BPSysAve` for consensus systolic blood pressure measurement.

- Identify **one or two** hypotheses for each aim.
  - *Don't* seek to analyse every variable in the dataset.

- **Not** expected or encouraged to do additional research about risk factors of high blood pressure.
  - Focus effort on defining testable hypotheses and clearly reporting statistical analyses to test.

Today

- Specify research questions for Aims 1 and 2.
  - Identify the **target population**.
  - Identify 2-3 statistical hypothesis that you will test

After 45 minutes: Present 1 PowerPoint slide stating your target population and hypotheses.

- Define dataset: any inclusion or exclusion criteria
- Identify variables that you will use: primary outcome(s), main exposure(s), covariates.
  - Quality of measurement and missingness
- Agree statistical methods
- Develop shell tables

15 minutes before session end: Present 1 PowerPoint slide identifying your outcome and exposure variables for each hypothesis.

Applied Stats Lab: Room Assignments

	G64				G65		
Group 1	Jingxian Huang Elena Venero Garcia Hailey Gu Jian Chen Seth Howes	Group 4	Yuchen Xie Elin Rowlands Daniel Adams Lea Maria Khoueiry Vaishnavi Shridar	Group 7	Xihao Cao Thomas Allwright Emily Knight Mehak Gurnani Xheni Prebibaj	Group 10	Ciara Hamilton Alia Rafiq Huike Cheng Ria Sachdeva Helena Bicanic-Popovic
Group 2	Yiyang Shi Emmanuelle Kern Anu Bode-Favours Ka Ki Lui Siwei Wu	Group 5	Bing Chen Oliver Simmons Daniel Huntley Marina Berger Wenjia Zhang	Group 8	Chiara Pligersdorffer Angela Aumonier Fiona Rice Nicole Cizauskas Yang Shen	Group 11	Sandra Gudziunaite Aditya Ramani Jaidip Gill Robert Campbell Onyango Sangoro
Group 3	Shuhui Li Pin-Chun Wang Cameron Appel Kheerthiharan Saravanan Sreenidhi Venkatesh	Group 6	Mi Ma Harrison Goldspink David Ensor Megan Pete Wenqi Cho	Group 9	Mathias Brugel Abdul-Hakeem Khan Gabrielle Provost Omar Eweis Yuju Ahn	Group 12	Juliet Arukwe Gillian Sigle-Hall James Tait Samuel Quill