MRC/CSO Social and Public Health Sciences Unit




Measurement and Bias

Dr. Megan McMinn

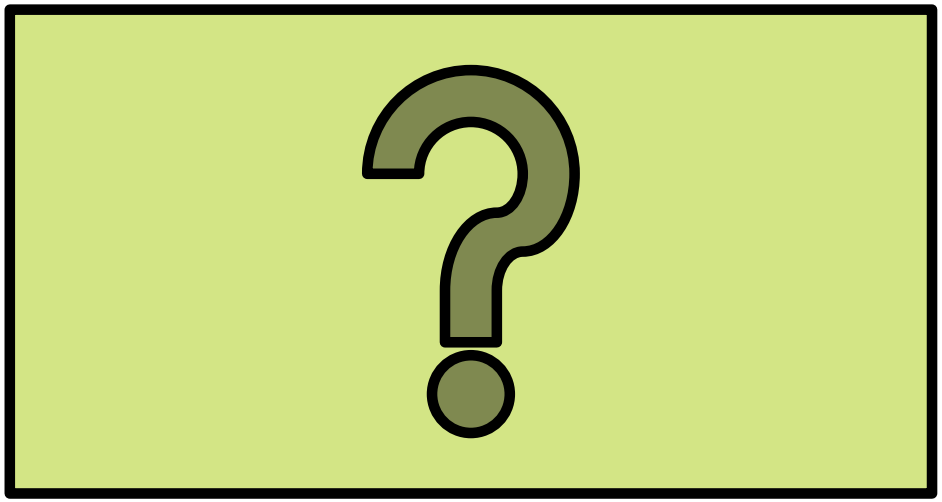
Megan.McMinn@glasgow.ac.uk

MRC/CSO Social and Public Health Sciences Unit, University of Glasgow

5 March 2019

Motivation

Why could measurement errors be more likely in epidemiology than in other disciplines?



Motivation

- There are some measures we can never know

- A given study can provide an estimate of it, but as we can't compare it to the true value, we cannot determine the actual amount of error in any given study.

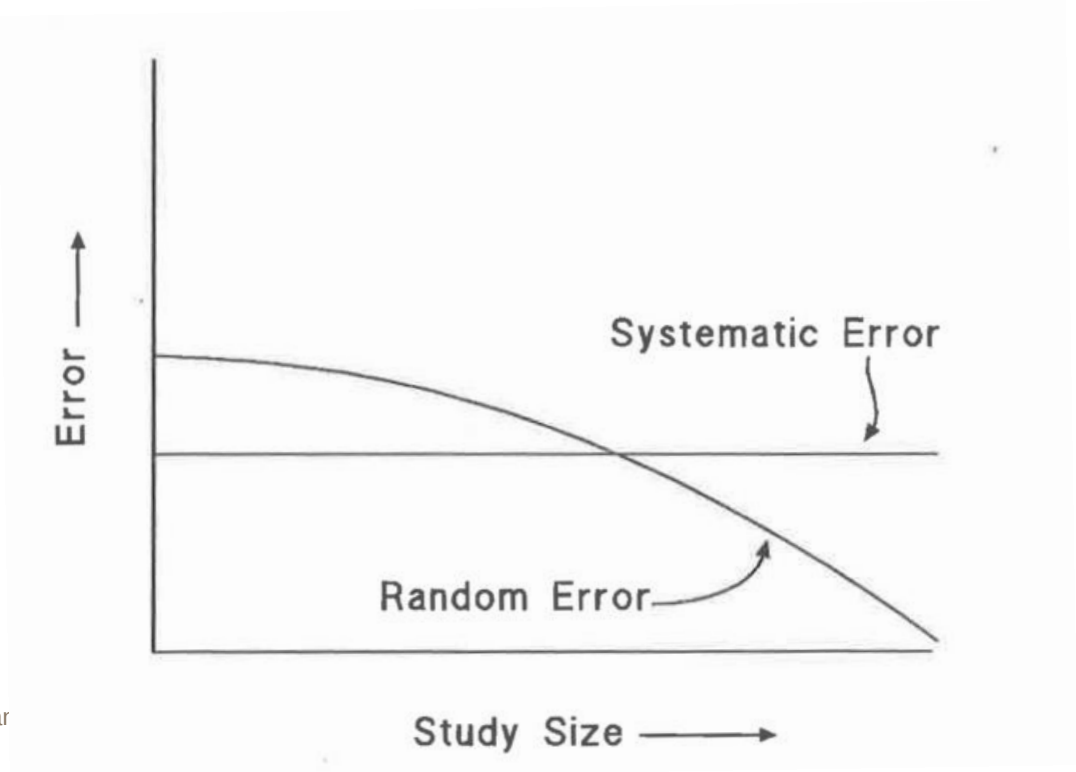
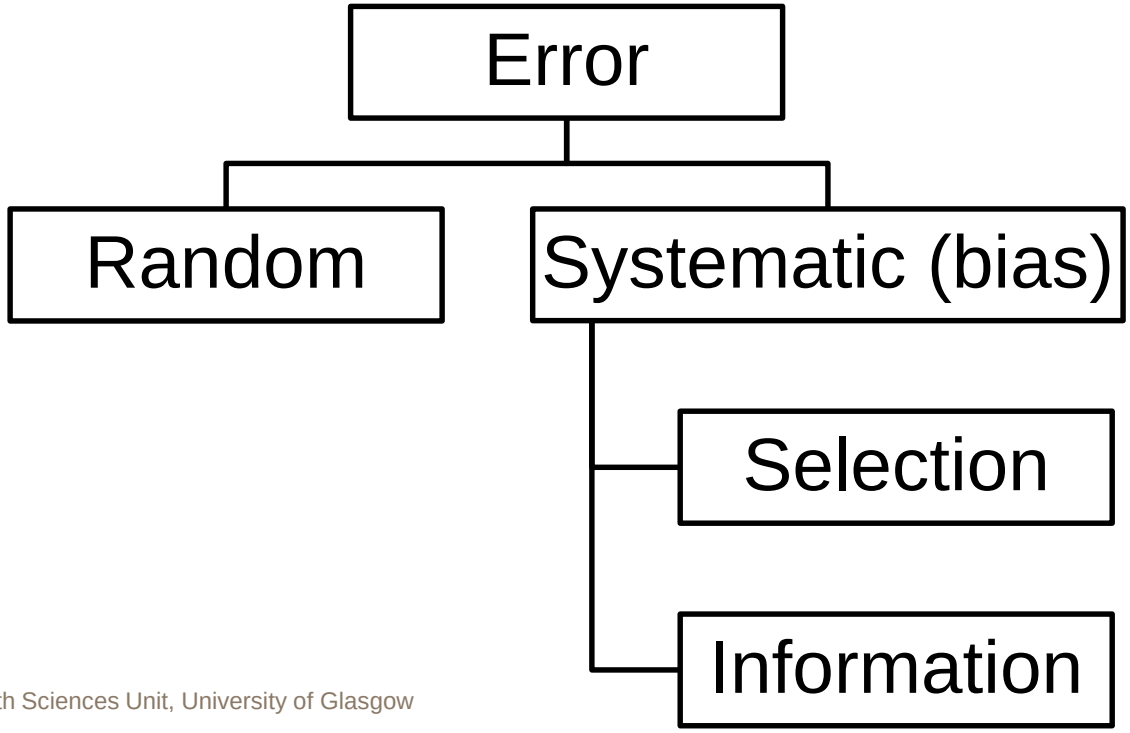


Fig 7.1 (Rothman 2012)

Aims/Learning Outcomes

Recognise important biases and understand:

- how these affect interpretation of findings
- how such biases can be dealt with through study design and/or statistical analyses.



Surveys

- Sample of the population at a specific point in time
- Used to measure prevalence of diseases and health behaviours, or opinions
  - Development, implementation and evaluation of social and public health policy
  - Important that the sample are representative of the population
    - Findings can be used to make claims about the whole population

Scottish Health Survey

- Commissioned by Scottish Government Health Directorate to provide reliable information on health, and factors related to health, of people living in Scotland that cannot be obtained from other sources
- 1995, 1998, 2003, annually since 2008
- Questionnaire & measurements

Selection bias

- A systematic error in the selection of study participants such that the association between exposure and outcome is different for those who are studied that it is for those who would be eligible but are not studied.

  - Sampling bias
  - Volunteer bias
  - Disease spectrum bias
  - Referral bias
  - Non-response bias
  - Healthy worker bias
  - Survival bias
  - Attrition (loss to follow-up) bias

Selection bias

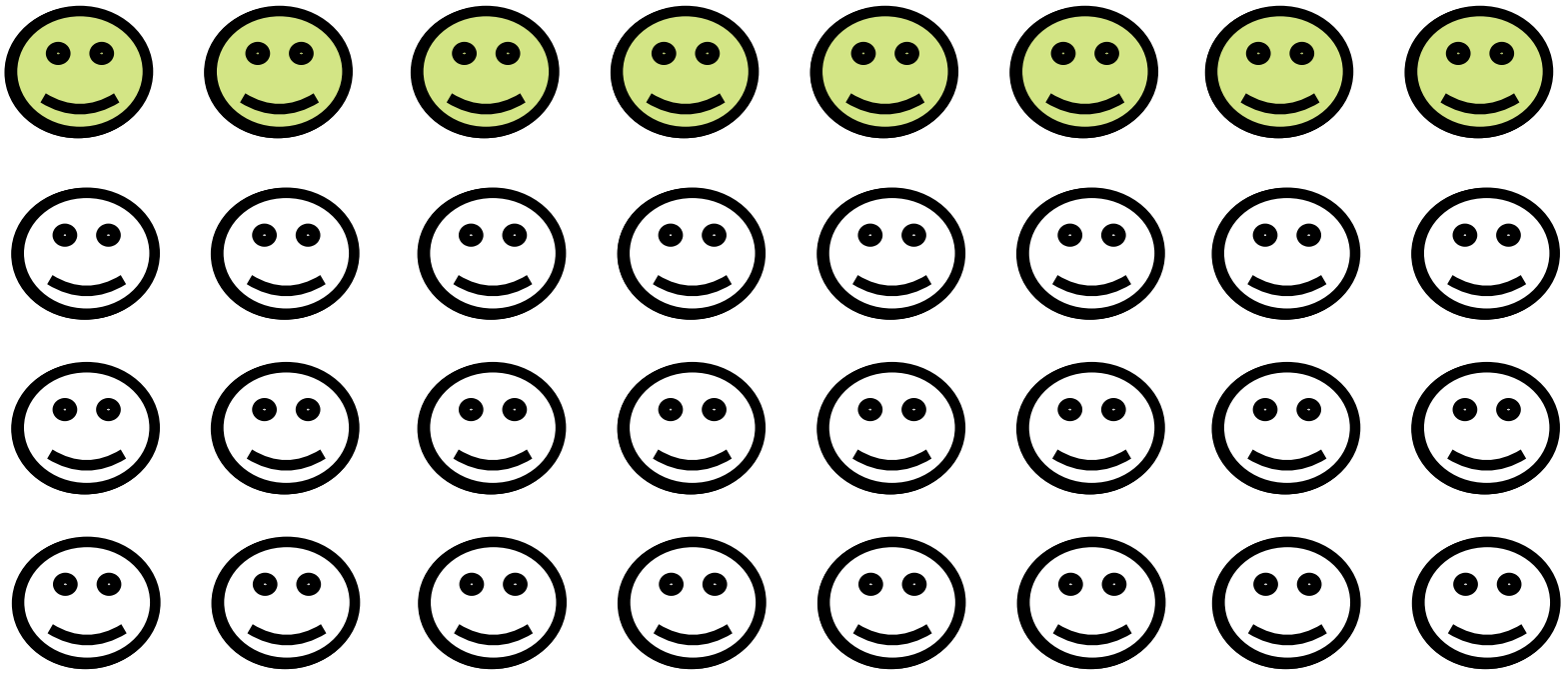
- Sampling – **random**, systematic, convenience, cluster, …

Selection bias

- Sampling – random, **systematic**, convenience, cluster, …

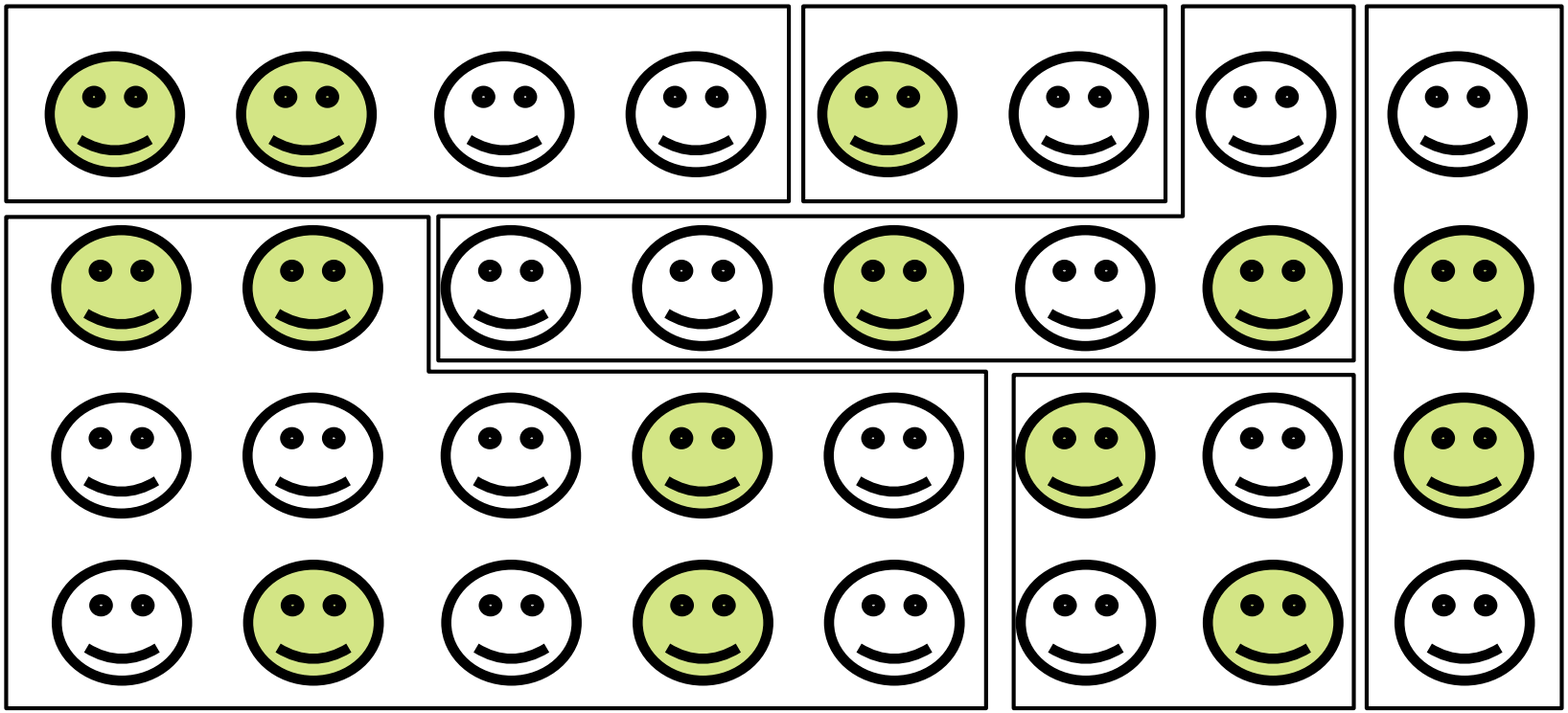
Selection bias

- Sampling – random, systematic, **convenience**, cluster, …



Selection bias

- Sampling – random, systematic, convenience, **cluster**, …



Selection bias

- Sampling
  - Think about the sampling frame - who could it be missing?

- Telephone directory
  - anyone without a landline, or those with landlines, but are ex-directory
- GP registration
  - anyone not registered with a GP – tends to be young men
- Scottish Health Survey - database of all private residences in Scotland
  - Anyone living in an institution (care home, prison)
  - Homeless

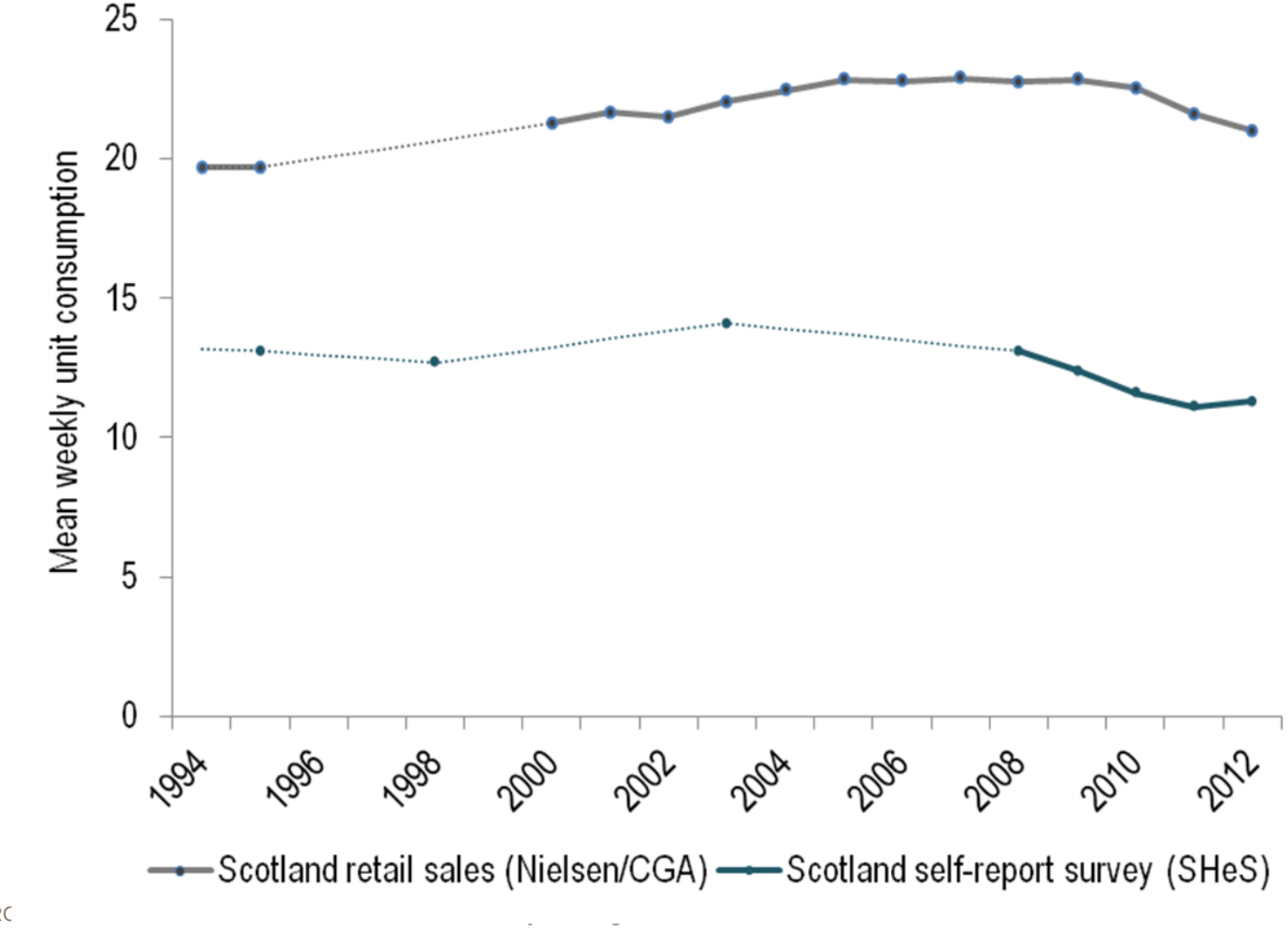
Selection bias

- Volunteer – tend to be different in their attitudes, behaviours and health status
- Non-response – those who refuse to participate/can't be contacted may be different from participants in terms of attitudes, behaviours and health status

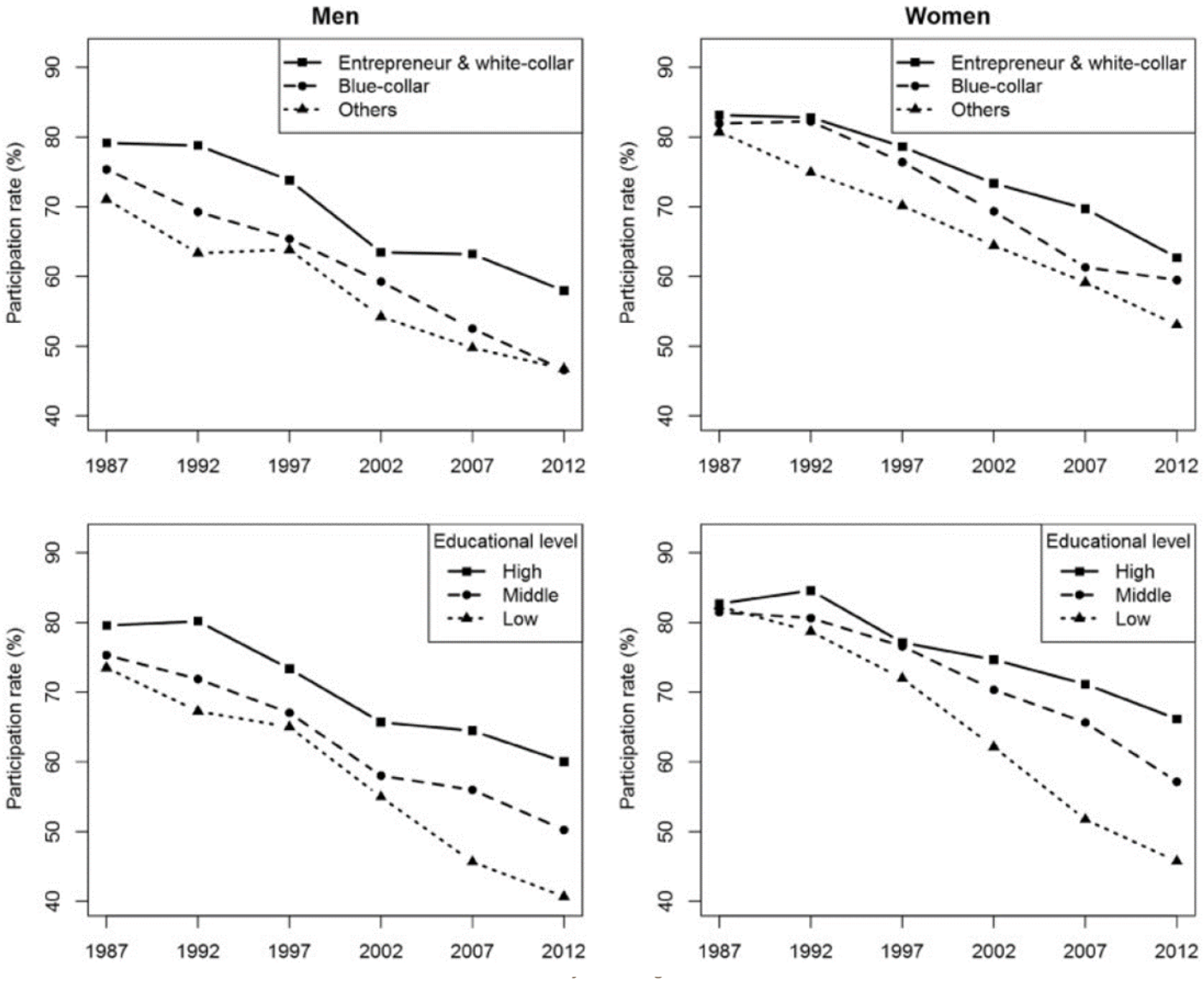
Non-participation bias

- Declining levels of participation cause:
  - Loss of information
  - Validity of study results are threatened if the participants and the non-participants differ systematically
    - Difficult in many settings as we don’t know who the non-participants are

Alcohol consumption in Scotland



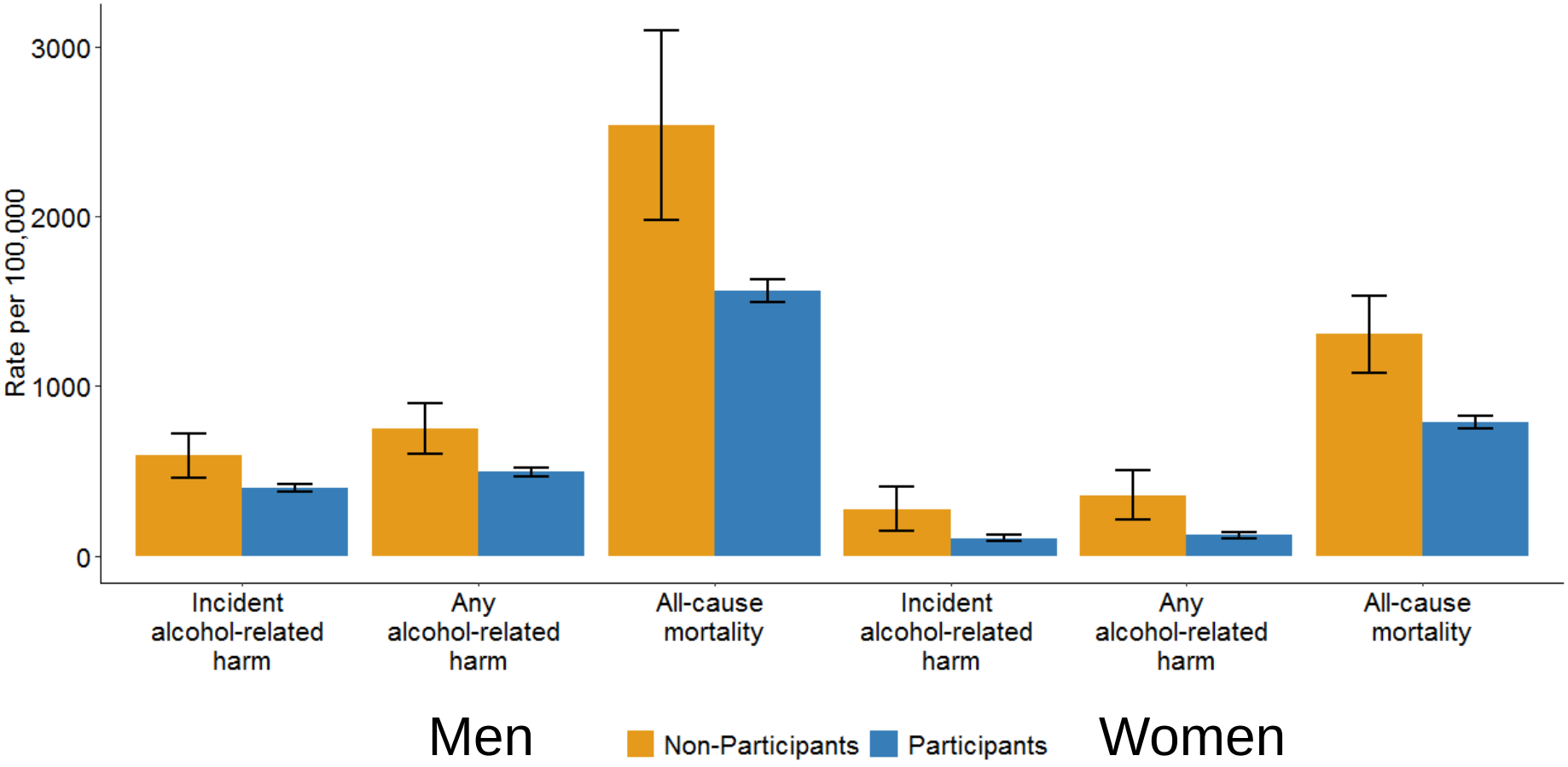
Non-participation bias in Finland



REINIKAINEN, J., et. al. 2018. Participation rates by educational levels have diverged during 25 years in Finnish health examination surveys. *European Journal of Public Health,* 28**,** 237-243.

Non-participation bias in Finland

See R Shiny
https://ihw-hehta.shinyapps.io/participationBias/



McMinn et al. (under review)

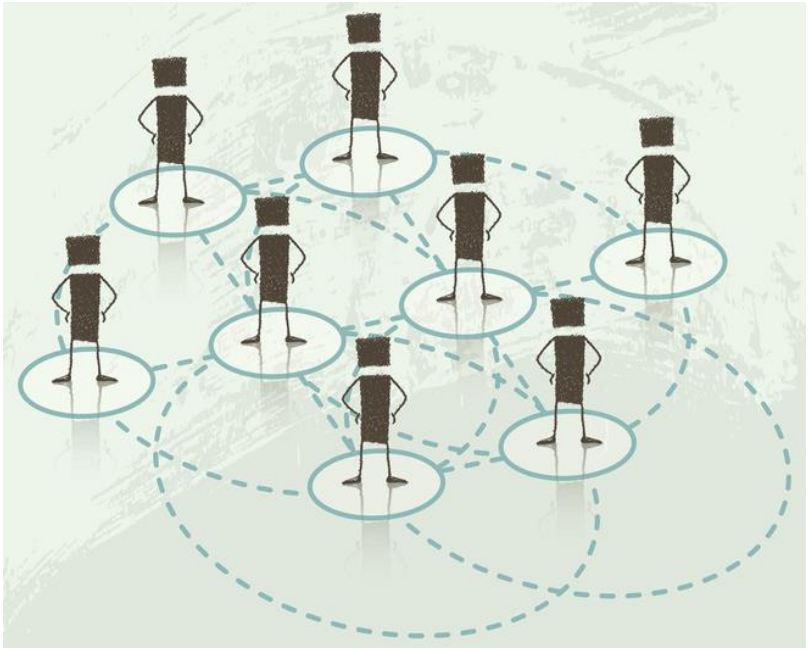
Selection bias

- Referral – individual preferences/local practices determine which subjects are referred for further investigations
- Healthy worker – ill people may be too ill to attend work, therefore sample of those at work is likely to be healthier
- Survival – sample population only made up of “survivors”

Representativeness

- Ideally, we'd be able to study everyone in a population, in order to get an accurate picture of the population

- In practice – often not possible
- Samples drawn from population enable generalisations to be made about the population of interest.



Representativeness

- Weighting can be used to adjust the sample to be more representative of the target population

1. Design weights – adjusting for sampling design
2. Non-response weights – adjusting to compensate for non-participants
3. Calibration/auxiliary data
4. Analyse weight variability

Internal and External Validity

Internal:

The degree to which the study correctly answers the research question

- How likely are the study results to be true?
- Increased by using randomisation, blinding, suitable inclusion/exclusion criteria, detailed measurement methods, appropriate statistical tests, promoting compliance

External:

Are the study results broadly applicable to more general groups of people/more realistic environments?

- Can these results be applied to other settings/populations?
- Increased by population wide recruitment, broad inclusion criteria, monitoring

Internal and External Validity

- Not all studies have to be externally valid
- A study which is internally valid but not generalizable is at least useful in informing practice in that setting
    - Snow's natural experiment of water pumps in London
    - Doll & Hill – mortality of male British doctors and smoking

- Little point trying to generalise an invalid study

Information/Measurement bias

- Distortion in measures of association caused by inaccurate measures of study variables

- The gold standard procedures for acquiring human study data are typically very expensive, intrusive, and impractical to conduct in large studies
  - More practical procedures may be used, and are generally fairly accurate, but there may be some error in the measures

Sources of measurement bias

- Interviewer bias
  - Lack of equal probing between outcome groups
  - Lack of equal measurement of outcome between exposure groups
- Subject self report
  - Incorrect recall
  - Reluctant to be truthful
- Procedure failure
  - Limited knowledge
  - Lab/equipment error
  - Records incorrectly coded

→ Misclassification

Types of Misclassification

Non-differential misclassification

- Misclassification is not related to exposure or disease status
- Probability of incorrect information is equal in the two group

Differential (systematic) misclassification

- Misclassification results in incorrect exposure being recorded in more cases than controls, or vice versa
- One group has more incorrect information than the other

Example I

Association of heavy alcohol use with depression

- Hypothetical prospective study - cohort of 1000 students
- Free of major depression at baseline
- Measurement – gold standards vs practical measurements

	Outcome	No outcome	Total
Exposed	30	220	250
Not exposed	15	235	250

Relative Risk = $\frac{30/250}{15/250} = 2.0$

Non-differential Misclassification

See R Shiny
https://ihw-hehta.shinyapps.io/misclassifcationExposure/

Association of heavy alcohol use with depression

- Hypothetical prospective study - cohort of 1000 students
- Free of major depression at baseline
- Measurement – gold standards vs practical measurements
- 20% misclassified exposure

	Outcome	No outcome	Total
Exposed	~~30~~ 24	~~220~~ 176	~~250~~ 200
Not exposed	~~15~~ 21	~~235~~ 279	~~250~~ 300

$$\text{Relative Risk} = \frac{24/200}{21/300} = 1.71$$

Example I

Association of heavy alcohol use with depression

- Hypothetical prospective study - cohort of 1000 students
- Free of major depression at baseline
- Measurement – gold standards vs practical measurements

	Outcome	No outcome	Total
Exposed	30	220	250
Not exposed	15	235	250

Relative Risk = $\frac{30/250}{15/250} = 2.0$

Non-differential Misclassification

Association of heavy alcohol use with depression

- Hypothetical prospective study - cohort of 1000 students
- Free of major depression at baseline
- Measurement – gold standards vs practical measurements
- 10% misclassified outcome

	Outcome	No outcome	Total
Exposed	~~30~~ 52	~~220~~ 198	250
Not exposed	~~15~~ 39	~~235~~ 211	250

Relative Risk = $\frac{52/250}{39/250} = 1.35$

Example II

Physical exercise and risk of miscarriage

- Cohort study of ~ 92,000 pregnant women
- Exposure - physical activity - obtained via interviews at 12-16 weeks gestation.
- Outcome – miscarriage - obtained from national registers.

Madsen, M. et al. (2007), Leisure time physical exercise during pregnancy and the risk of miscarriage: a study within the Danish National Birth Cohort. BJOG: An International Journal of Obstetrics & Gynaecology, 114: 1419-1426. doi:10.1111/j.1471-0528.2007.01496.x

Findings:

"Certain types of exercise, and particularly high impact types of exercise, were found to be **associated with a higher risk of miscarriage**"

After excluding women who had experienced the outcome before being interviewed, there was **no evidence of an association**.

Example II

Potential misclassifications:

Of the outcome?

- unlikely, outcomes obtained from health records

Of the exposure?

- How is the exposure measured?
  - Self-reported physical activity → recall bias
  - Non-differential?
  - Differential?

Differential misclassification

- Misclassification errors in exposure measurements differ between those with and without the outcome

- In Example II - Physical activity in women who had experienced the outcome (miscarriage) was over-reported relative to those without the outcome
  - Note – does not mean that the retrospective reported exposures were less valid

Kestenbaum (2018)

Example III

- Misclassification errors in outcome measures differ systematically between exposure groups

	Outcome	No outcome	Total
Treatment A	14	11	25
Treatment B	15	60	75

Relative Risk = $\frac{14/25}{15/75} = 2.8$

Kestenbaum (2018)

Differential misclassification

- Misclassification errors in outcome measures differ systematically between exposure groups
- Implement blinding

	Outcome	No outcome	Total
Treatment A	~~14~~ 10	~~85~~ 15	25
Treatment B	15	60	75

$$\text{Relative Risk} = \frac{10/25}{15/75} = 2.0$$

Kestenbaum (2018)

Non-differential Misclassification

- Arises from non-systematic errors in measuring study data
- … of the exposure - Misclassification errors are assumed to be similar in each outcome group
- … of the outcome – misclassification errors are assumed to be similar across exposure groups

- Typically results in RR → 0 (bias towards null)
  - Association may weaken, or become statistically non-significant

Differential misclassification

- Arises from systematic errors in measuring study data
- … of the exposure – misclassification errors are different between outcome groups
- … of the outcome – misclassification errors are different between exposure groups

- Can result in stronger, weaker, or false associations
  - Study measurements need careful consideration
  - Has the ability to undermine studies

Misclassification in research

- Data collection methods described in Methods section
- Subjective – can't go back and repeat the study using gold standard measures

1. What measures could be misclassified (outcome, exposure, other data?)
2. Is the suspected misclassification differential or non-differential?
3. What is the expected impact of this misclassification?

Minimising bias in data collection

- Ensure study population are representative of the target population
  - Think about sampling frame – who will it be missing?
- Ensure a high response rate and minimal loss to follow-up
- Clearly define exposure and outcome
- Clear protocol for collecting, recording, coding and cleaning data
- Reliable and valid measurement methods
- Blinding

Further reading/tools

- Understanding Health Research Tool - Understandinghealthresearch.org

- Rothman, K. J. 2012. *Epidemiology: an introduction,* Oxford; New York, N.Y, Oxford University Press.
- Kestenbaum, B., 2018. *Epidemiology and biostatistics: an introduction to clinical research*. Springer.
- Bhopal, R.S., 2002. *Concepts of Epidemiology: An integrated introduction to the ideas, theories, principles and methods of epidemiology.* Oxford University Press.

From Intro to Epi:

- Delgado-Rodriguez, M. & Llorca.J. “Bias”, J Epidemiol Community Health 2004;58:635–641. doi: 10.1136/jech.2003.008466
- Sica, G. “Bias in Research Studies”, ***Radiology:*** Volume 238: Number 3 (2006)

Break

Exercise

Read the Methods sections of two papers

Think about the topics we've covered today:

- Are the samples used representative?
- Could the results be generalised? (external validity)
- What are the outcomes/exposures of interest?
  - How are these measured?
  - Potential for misclassification?
- What biases could the papers be at risk of?

Paper 1

JAMA Network | Open

Original Investigation | Nutrition, Obesity, and Exercise

Exploring the Role of Family Functioning in the Association Between Frequency of Family Dinners and Dietary Intake Among Adolescents and Young Adults

Kathryn Walton, PhD, RD; Nicholas J. Horton, ScD; Sheryl L. Rifas-Shiman, MPH; Alison E. Field, ScD; S. Bryn Austin, ScD; Emma Haycraft, PhD; Andrea Breen, PhD; Jess Haines, PhD, RD

Accessed - https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2715616 (28/02/2019)

Table 1. Participant Characteristics According to Sex and Family Dinner Frequency

Characteristic	Total (N = 2728)	Adolescents and Young Adults		Family Dinner, Times/wk	
		Male (n = 1169)	Female (n = 1559)	<5 (n = 1681)	≥5 (n = 1047)
Age, mean (SD), y	19.4 (1.9)	19.4 (1.9)	19.5 (1.9)	19.7 (1.9)	19.0 (1.8)
Race, No/total No. (%)					
White	2453/2649 (92.6)	1050/1145 (91.7)	1403/1504 (93.3)	1511/1629 (92.8)	942/1020 (92.4)
Other	196/2649 (7.4)	95/1145 (8.3)	101/1504 (6.7)	118/1629 (7.2)	78/1020 (7.6)
Family structure, No./total No. (%)					
Dual-parent home	2155/2685 (80.3)	933/1160 (80.4)	1222/1525 (80.1)	1308/1659 (78.8)	847/1026 (82.6)
Single-parent home	530/2685 (19.7)	227/1160 (19.6)	303/1525 (19.9)	351/1659 (21.2)	179/1026 (17.4)
Mother's spouse or partner's educational level, No./total No. (%)					
<College education	806/2576 (31.3)	343/1105 (31.0)	463/1471 (31.5)	544/1570 (34.6)	262/1006 (26.0)
College education	1770/2576 (68.7)	762/1105 (69.0)	1008/1471 (68.5)	1026/1570 (65.4)	744/1006 (74.0)
Family functioning, No. (%)					
High[a]	2091 (76.6)	889 (76.0)	1202 (77.1)	1223 (72.8)	868 (82.9)
Low[a]	637 (23.4)	280 (24.0)	357 (22.9)	458 (27.2)	179 (17.1)
Family functioning score, mean (SD)[a]	1.9 (0.5)	1.9 (0.5)	1.8 (0.5)	1.9 (0.5)	1.8 (0.5)
Family dinner frequency, mean (SD), dinners/wk	3.4 (1.6)	3.4 (1.6)	3.3 (1.7)	2.4 (1.3)	5.0 (0.0)
Fruit without juice, mean (SD), servings/d	1.5 (1.6)	1.3 (1.4)	1.5 (1.8)	1.3 (1.5)	1.7 (1.8)
Vegetables, mean (SD), servings/d	2.3 (2.7)	2.1 (2.6)	2.4 (2.7)	2.1 (2.4)	2.7 (3.1)
Sugar-sweetened beverages, mean (SD), servings/d	0.8 (1.2)	1.1 (1.5)	0.5 (0.8)	0.8 (1.2)	0.7 (1.1)
Fast food, mean (SD), times/wk	1.0 (1.3)	1.2 (1.5)	0.8 (1.1)	1.1 (1.4)	0.8 (1.1)
Takeout food, mean (SD), times/wk	0.9 (1.1)	0.9 (1.1)	0.8 (1.0)	0.9 (1.2)	0.7 (0.9)

[a] Lower family functioning scores indicate better family functioning. A score less than 2.17 indicates high family functioning; 2.17 or higher indicates low family functioning.

Paper 2

Original Investigation

ONLINE FIRST

February 11, 2019

Association Between Ultraprocessed Food Consumption and Risk of Mortality Among Middle-aged Adults in France

Laure Schnabel, MD, MSc[1,2]; Emmanuelle Kesse-Guyot, PhD[1]; Benjamin Allès, PhD[1]; et al

» Author Affiliations | Article Information

JAMA Intern Med. Published online February 11, 2019. doi:10.1001/jamainternmed.2018.7289

Accessed - https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/2723626 (28/02/2019)

Exercise

- The Discussion section of each paper contains the Strengths and Limitations for the paper, where most of what we've discussed will be included.