INTHE4021 **INTRODUCTION TO EPIDEMIOLOGY**, Exam

# Research question and design

1. You are planning a research project, based on the following research questions: “How common is Diabetes Type 2 among adults in country x”; and “Is there an association between physical inactivity and Diabetes Type 2 among adults in country x”.

a) Give two specific objectives that can be achieved in order to give answers to the research questions?

The objectives of the present study among men and women aged 40-85 year olds in country x are:

1. to estimate the prevalence (alt. incidence) of diabetes type 2;
2. to estimate the association between physical inactivity and diabetes type 2.

b) Select a research design and give pros and cons for the selected design. Explain why you did not choose another design?

Cross-sectional design; cohort design; RCT (discuss strengths and weaknesses of the different designs)

c) Describe the target population and sample, including the sample selection method.

Target pop.: 40-85 year old men and women in country x;

Sampling method: randomly selected clusters of three cities with surrounding rural areas; 1000 participants (based on power calculation) selected randomly proportionally to the sample size of the three clusters

d) Make a plan for the main analyses of the study.

Draw a DAG (use Dagitty), describe that you will estimate prevalence/incidence with 95% CI, and logistic regression analyses to estimate PR, OR or RR with 95% CI, and test for association using Chi-square test.

e) Describe potential selection problems (bias) and information problems (bias) in the study

Selection bias: Cross-sectional design: if several of the oldest refuse to participate (underestimation of prevalence); if several of those with both physical inactivity and diabetes type 2 refuse to participate, there will be an underestimation of the association between physical inactivity and diabetes type 2.

OR Cohort design: if large loss to follow-up among the oldest (underestimation of incidence of diabetes). If large loss to follow-up among those who are physically inactive (who are likely to develop diabetes type 2), there will be an underestimation of the association between physical inactivity and diabetes type 2.

Information bias: systematically erogenous measure of exposure or outcome leading to over- or underestimation of exposure and/or outcome

# Measures



Calculate Risks and Odds for having diabetes among Inactive and Active based on the table above.

1. What is the risk of diabetes among the inactive? 36%
2. What is the risk of diabetes among the active? 6%

Calculate the Risk Ratio and The Odds Ratio for the effect of inactivity on Diabetes. Choose active as the reference group.

1. What is the Risk Ratio? RR=5.8
2. Is the Risk Ratio similar to the Odds Ratio? OR=8.5, larger than RR since diab is common

#  DAGs



1. We want the total effect of Inactivity on Diabetes, what should we adjust for to remove confounding based on the DAG above? Age, sex and education
2. To get the direct effect of inactivity, we adjust for age, sex, education and BMI. What would happen? We would not get the direct effect of inactivity, because a collider path would open
3. To get the direct effect of inactivity, we adjust for age, sex, education, BMI and diet. What would happen? We would get the direct effect of inactivity

# Regression



Above are descriptive tables of blood glucose and diabetes type 2. We want to do a regression with glucose as the outcome, and a regression with diabetes as the outcome.

A) How do we choose what type of regressions (linear/logistic) to use? We typically use linear regression for continuous outcomes, and logistic regression for binary outcomes.

# Linear Regression





We have listed the label for inactivity and done a linear regression with glucose (gluc) as the outcome and inactivity (inact) as the exposure.

A) What interpretation of the result is correct? Inactive have 2.02 units higher glucose than active. The effect is significant because the p-value is less than .05



Above is a table comparing the unadjusted linear regression model (m1) with the adjusted model (full).

B) What is the unconfounded effect estimate of physical inactivity on blood glucose? 1.24

# Logistic Regression



Above is the logistic regression with diabetes (diab) as the outcome and physical inactivity (inact) as the exposure

A) What is the correct interpretation of the result? The inactive have 8.5 times higher odds of diabetes than the active. The effect is significant.



 The table above compares the unadjusted logistic model (m1) with the adjusted model (full).

B) What is an unconfounded estimate of the OR of inactivity on diabetes? OR=1.9