Solutions 3: Ecologic and Cross-sectional studies

Question 1: Circumcision and HIV

Question 1a: Exposure: male circumcision; measure: prevalence
Outcome: HIV seropositivity; measure: prevalence
Study populations: national populations in sub-Saharan Africa.

Question 1b: Sub-Saharan countries with lower levels of male circumcision generally have higher HIV prevalence than countries with higher levels of male circumcision.

Question 1c (I): Study design: Ecologic (geographic)

Potential data sources: HIV status from nationally representative sources (in practice might not be national), such as:

- Demographic and Health Surveys (DHS) (not routine but frequent in many countries)
- National Health Surveys (not routine but frequent in many countries)
- Blood tests among antenatal care users (routine facility data)
- Surveillance of high-risk groups (routine sentinel surveillance)
- Unlinked anonymous studies (routine in blood banks etc)

Estimates of male circumcision are now obtained from special surveys (e.g. DHS) which are not routine. Earlier estimates were obtained from ethnographic sources.

Question 1c (II):

- Data on the prevalence of HIV may not be nationally representative (depends on source, but for example may be from a sub-national sample frame).
- These data are almost certainly from tests (rather than self-report) but if they were obtained through self-report then they are likely to be very biased as people may not know status, may miss-report due to stigma, etc.
- Tests also may not be entirely accurate (e.g. may be screening rather than confirmatory).
- Male circumcision status assessment may also not be in nationally representative sample (again, depends on data source). In some countries overall prevalence of male circumcision is very low but certain communities may practice it routinely - so if survey is only among select communities, it may not be nationally representative.
- Circumcision status may be reported not observed in some countries. Men may not know, and misreporting is common. Reporting is also complicated by partial male circumcision.
- Reports may be indirect (e.g. from women reporting on their partner); women also may not know.

In addition to the extent to which the measures of exposure and outcome are valid, the timing of data collection is also often of critical importance – for male circumcision

this is less likely to have changed dramatically over time if cultural practice, but may be an issue for HIV status. For example, Uganda had a higher HIV+ prevalence before 2005 (the point shown on the graph. Age of circumcision may be neonatal, but in some settings may occur after sexual debut (first sexual intercourse).

Question 1d:
Strengths: can use secondary (existing) data, may be quick, can be used to generate hypotheses.

Limitations: these are population-level, not individual-level associations. Uncircumcised men may not be more likely to be HIV+ within a country, even if there are associations at the national level. This would be an example of the "ecological fallacy".

There may also be confounding by other factors (e.g. male circumcision as a cultural practice, linked to religion and ethnicity). Circumcision may be associated with other exposures such as sexual behaviour (e.g. numbers of sexual partners) which may in fact be the true protective factors. You will learn about confounding soon.

Question 1e*:* In some countries male circumcision is performed in neonates and infants, and in this situation must precede sexual transmission of HIV (though not vertical transmission). Therefore, male circumcision is unlikely to be undertaken in *response* to seroconversion in most cases. Consequently, if an association is found, it is almost certain that male circumcision affects HIV status and not that HIV status affects male circumcision. Having said that, vertical transmission of HIV would be an exception to the exposure preceding outcome.

Question 2: The British National Surveys of Sexual Attitudes and Lifestyles

Question 2a: This was a cross-sectional study

Question 2b: The source population was all men and women living in Great Britain and aged 16 to 74 at the time of the survey (2010-2012).

Question 2c: The concern about response rates comes from the fact that non-responders may be different from responders in the characteristics you are measuring. For example, those who refuse to take part in the study may be more, or less, sexually active than the responders. Thus, results from the study (which be definition is obtained only from the responders) may not be representative of the general population of people aged 16-74. This is a form of selection bias, and you will hear more about this later in the course.

With regards what is a good response rate, there is no 'correct' answer to this. A response rate of 58% is not atypical for this sort of survey particularly because it was dealing with a very sensitive issue (sexual behaviour). Ideally, piloting a survey would be done beforehand to see what the likely response rate would be, and to identify any factors that could help increase the response rate.

The key thing to remember is that reasonably valid answers can be obtained from a study with a 58% response rate if it is designed and conducted well, and if the non-responders do not differ greatly from those who do respond. Conversely, a (very) high response rate may simply be an artefact of a poorly designed and conducted study! However, there is an emerging literature to suggest that a low response rate may not compromise the external validity of the data [1]. However, in many cases we just do not know whether responders and non-responders do differ in ways that may bias the study.

Of course, there is a problem here because you do not have any information on the outcomes of interest for the non-responders, so how can you tell if they are different from the responders? What is normally done is a comparison of the information you do have, such as age and sex distributions, for the responders and non-responders. If the distributions of these variables are very different, you have evidence that the results of your study will not be representative of the general population. Another thing you can do is compare data from the responders, such as age, sex and other socio-demographic variables, with that obtained from other well-conducted population-based surveys.

You will learn more about these concepts in teaching week 4 (chance & bias).

Question 2d: The NATSAL-3 questionnaire was extensive, and had sections on:

- health, family and learning about sex
- first sexual experiences, use of contraception and sexual lifestyle
- number of sex partners, sexual practices, sexual health
- attitudes and risks
- socio-demographic questions

Some examples of the specific measures:

- Had sex in the past year (period prevalence)
- Tested for HIV in past 5 years (period prevalence)
- Current infection with HPV, Chlamydia, Gonorrhoea, or HIV (prevalence)

Read more here: http://www.natsal.ac.uk/natsal-3.aspx

Question 3: Deciding what was done

Question 3a

STUDY DESIGN?	Ecologic study (geographic)
SOURCE POPULATION	Infants in the 15 study countries in 1990
PRIMARY OUTCOME?	Infant mortality
PRIMARY EXPOSURE?	Air pollution
MEASURE OF DISEASE (OUTCOME) OCCURRENCE?	What is often calculated is infant deaths in 1990 per 1000 live births in 1990. This is commonly described as a ‘rate’.

	In fact it is closest to an infant mortality risk (i.e. probability of infant death among births in 1990). NB some of these deaths may occur after 1990, say for a baby born in December 1990, but is offset by infant deaths in 1990 from babies born in 1989
POSSIBLE APPROACHES/MEASURES TO ASSESS THE EXPOSURE-OUTCOME RELATIONSHIP?	Average change in infant mortality per unit change in air pollution (i.e. the slope of a linear regression)

Question 3b

STUDY DESIGN?	Cross-sectional study (analytic)
SOURCE POPULATION	All individuals in the community studied
PRIMARY OUTCOME?	Trachoma (Yes/No)
PRIMARY EXPOSURE?	Crowding in housing; possibly measured as a binary indicator (e.g. one room for each single person aged 18+ years? Y/N) or a continuous variable (e.g., no. household members per km^2)
MEASURE OF DISEASE (OUTCOME) OCCURRENCE?	Prevalence of trachoma
POSSIBLE APPROACHES/MEASURES TO ASSESS THE EXPOSURE-OUTCOME RELATIONSHIP (consider only for a binary measure of exposure)?	For binary exposure: prevalence ratio, Prevalence difference, Prevalence odds ratio between the two groups (exposed and unexposed to crowding in household)

Question 4

To do in your own time: Visit the Gapminder website and explore the various datasets to create graphs - instant ecological studies! Consider the quality of data for the chosen variables and what the relationship may (or may not) show.

References

1. Hendra R, Hill A. Rethinking Response Rates: New Evidence of Little Relationship Between Survey Response Rates and Nonresponse Bias. Eval Rev 2019; 43(5): 307-30.