FASM1 – 1er semestre 2020 : Session 2

The cohort study : part 1

In today's session we are going to look at a type of research article you may already have come across.

History

As early as 1912 British doctor and researcher, Janet Lane-Claypon had been experimenting with the use of cohorts in studies on breast milk, with a view to studying the relationship between breastfeeding and growth in babies. In 1935 US epidemiologist Frost first used the term cohort study. Since then, like most study types, it has become a standardised design.

A cohort study uses a very large population which is followed over a long period to study one or more outcomes. Data can be collected from different sources such as questionnaires, interviews, consultations, medical tests or biological samples, or medical records. Data can be examined prospectively or retrospectively.

There is a considerable number of existing population cohorts. The UK medical research council lists almost 50 in the British Isles alone. Often set up to deal with a specific question in a specific category of the population, a cohort usually contains several thousand participants. Some examples of major cohorts are the *US Nurses Health Study* whose original aim was to study the consequences of oral contraceptives, or the *Framingham Heart Study* which is focussed on cardiovascular risk. In Denmark the entire population has become a cohort for research. In France there is a 200 000-person cohort called *Constances*.

While most cohorts begin with a specific aim, and the data collected is designed for prospective use, it is frequent for a cohort's data to be used retrospectively, and also for studies to "piggyback" on a relevant cohort, in some cases allowing for collection of additional data aimed to answer a specific question which was not anticipated initially.

In the case of a prospective cohort study, participants are not initially affected by the outcome of interest. Potential risk factors will be measured at baseline and at regular intervals over the course of the period initially planned.

The data analysis type most frequently used in cohort studies is relative risk. This is calculated by dividing the incidence rate of those exposed to the risk factor, by the incidence rate of those who are not exposed.

Multiple outcomes can be assessed, and multiple and rare risk factors can be observed. Exposure is assessed first in a prospective approach, thus enabling researchers to establish chronology and a causal relationship. Both incidence and prevalence can be assessed given the multiple data collection points.

However, this approach is costly, which explains why these population cohorts tend to be government funded initially, with the aim to study a major issue in public health. Any long-term study will suffer from attrition as people move house, die, or simply tire of participating. There is a high risk of potential confounders, so researchers should take this into account in their data analysis. There is also potential risk of observer bias, or behavioural modifications among participants, depending on the mode of data collection. Furthermore, should diagnosis change, as happened with Autism Spectrum Disorder in between versions IV and V of The Diagnostic and Statistical Manual of Mental Disorders (DSM) this can change the status of participants.

FASM1 – 1er semestre 2020 : Session 2

Structure

You are undoubtedly already familiar with the IMRaD structure, used by most original research articles. This acronym describes the structure of the body text: Introduction, Methods, Results and Discussion.

However, an article is not merely composed of these four sections, and it contains *paratextual elements* both before and after the body. These include the title, abstract and references, and all play a vital role in communicating the authors' work.

Each part of the article from the title, all the way through to the references has a pragmatic function, and this will influence its structure and characteristics.

For example, the title is the first element the reader encounters, and will therefore serve as a deciding element in the question as to whether to continue reading or not. It therefore needs to fulfil its main function, which is not to attract, but to inform, and the structure and language of the title should therefore reflect this aim.

Today's task

1. Observe the article you have downloaded

   Li, Y., Schoufour, J., Wang, D.D., Dhana, K., Pan, A., Liu, X., Song, M., Liu, G., Shin, H.J., Sun, Q. and Al-Shaar, L., 2020. Healthy lifestyle and life expectancy free of cancer, cardiovascular disease, and type 2 diabetes: prospective cohort study. *bmj*, *368*.

2. Identify all the parts which precede and follow the main body, and then make a complete list (including the main IMRaD parts). Then note what you think the function of each section is, taking notes on both the internal structure (can you identify how it is constructed?) and the linguistic features (for example, the tenses that are used) for each part.

3. Read the article in full before the next session

You are of course welcome to work in twos or threes, and you may contact me with any questions.

A document with a completed table will be available on my website within 48 hours.