

International Journal of Population Data Science

Journal Website: www.ijpds.org



Using SQL for accessing and working with large administrative data in TREs

Tudor Vilcan¹

¹Office for National Statistics, Titchfield, United Kingdom

Objectives

The objective of this presentation is two-fold. First, it will explain why SQL is needed for large admin data provision and how this is operationalised in practice in TREs. Second, it will provide researchers with guidance and instructions for manipulating, cleaning and analysing data in SQL.

Methods

Administrative datasets such as the invaluable Longitudinal Education Outcomes (LEO) are characterized by very large size and number of variables, as well as deep row counts. This entails it is not suitable to be provisioned via means such as flat files and software more familiar to researchers, such as STATA, SPSS or R. SQL is a valuable and sometimes indispensable tool for the provision of such data, as it can provide adequate storage, selective access to prevent disclosure risk and tools for data manipulation. The presentation will also cover weaknesses of SQL and when other software is more appropriate to use.

Results

Through this presentation, we hope to ensure researchers will have a much better understanding of how and why data is provisioned through SQL, especially large administrative datasets. They will also learn techniques and principles of SQL usage, alongside useful pieces of code and syntax. In addition, the presentation will highlight when SQL should be used and when it is time to employ a different software, as it has its distinct strengths and weaknesses.

Conclusion

Through understanding the use of SQL to access, manipulate and analyse large administrative datasets in TREs, researchers will be able to maximise the potential of their research and have a better experience overall.



Temporary page!

$\LaTeX$ was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because $\LaTeX$ now knows how many pages to expect for this document.