International Journal of Population Data Science

Journal Website: www.ijpds.org

Exploratory versus experimental design: overcoming the prejudice of 'data dredging'.

Pascoe, P^{1*} and Jones, SJ²

¹The Health Foundation

²Birmingham City University

As part of its most recent Efficiency Research Programme, which sought to advance understanding of workforce retention in health and social care, the Health Foundation have provided funding to a team within the Centre for Social Care, Health and Related Research at Birmingham City University. This project proposes to utilise exploratory data analysis techniques to investigate the underpinning contributory factors to nurse retention and ultimately patient safety across 10 acute and 10 mental healthcare providers. Exploratory data analysis techniques do not conform to the accepted hypothesis setting and testing associated with the traditional experimental research design within medicine and consequently, can be subject to criticism. However, these techniques afford bountiful opportunity to generate insight from real-world data; insight that could not be gleamed from experimental designs, therefore they warrant further investigation before being written off as 'data dredging'.

Exploratory data analysis techniques applied to large datasets that are routinely collected as part of standard care have the potential to revolutionise healthcare. Machine learning and artificial intelligence are becoming common place in everyday life but have been relatively slow to be adopted as a mechanism for improving healthcare. An aspect of the medical community's reluctance to accept these techniques is a perceived lack of robustness in determining relationships, as hypotheses aren't typically generated in advance and therefore the process exposes itself to unscrupulous actions where analyses are undertaken until a relationship is found. An element of this programme of work is to develop a set of standards for the application of these techniques to healthcare-generated data including, but not limited to and housing data in open access formats to improve transparency and reporting on findings.

This session seeks to explore these issues with colleagues across disciplines, generating debate regarding novel approaches to complex issues.

*Corresponding Author:

Email Address: polly.pascoe@health.org.uk (P Pascoe)