

Felix Ritchie (University of the West of England)

Spontaneous recognition: risk or distraction?

Topic 4 – Getting the statistics out

Keywords: confidentiality, privacy, statistical disclosure control, spontaneous recognition, data access

Introduction

'Spontaneous recognition' occurs when a microdata user believes that he or she has identified one of the respondents in the dataset: a neighbour, a co-worker, an organisation, or a group of patients, for example. The identification need not even be correct: the perceived breach of confidentiality can be as important as an actual breach.

This causes great concern to data providers wanting to allow researchers to use confidential data: no matter how trustworthy researchers are, they are still human and the spontaneous recognition of an individual might lead to the disclosure of information about an identified respondent. Hence, data providers are often insistent on minimising the risk of spontaneous recognition, and it creates an important hurdle to be addressed by those requesting access to data.

However, this paper argues that the hurdle is an unhelpful distraction, and other, more evidence-based, questions are more relevant.

Methods / Problem statement

Spontaneous recognition has little or no practical value. It is not possible to prove that a dataset has no chance of spontaneous recognition; we do not have access to all possible states of the world and information sets. It is difficult to conceive of circumstances under which one could show that spontaneous recognition by a researcher has occurred: how can we tell what is going on in someone's mind?

Finally, it is highly unlikely that there is any lawful or ethical basis for the concept: researchers have been given access to data under conditions of trust, and only if they make an unauthorised disclosure to a third party has that trust been breached. Nevertheless, most data providers adopt a defensive position, and so are likely to argue that the data should be protected against spontaneous recognition. In short, data are likely to be over-protected for no useful purpose.

Results / Proposed solution

We argue that a more relevant concept is 'identity confirmation': taking active steps to confirm a successful identification or assert one (again, whether the identification is accurate or not is not relevant). Unlike spontaneous recognition, this is not a statistical problem but an operational one, and so amenable to managerial solutions.

This issue raises important questions about the way data owners are persuaded to allow their data to be used. Of the two concepts (identity confirmation and spontaneous recognition), data providers should be concerned about identity confirmation, but they may be unable to articulate it as the literature focuses on spontaneous recognition as a statistical problem. Research data managers therefore have a role to play in making data providers aware of the specific risk being raised.

However, a more important problem is the default assumption that data providers require research managers to demonstrate that either concept is not a concern. As is well known, changing the default

assumption can have a major influence on data access decisions; in this case, the change in perspective focuses attention on the managerial issues.

We use access to business data as an example, as this is where spontaneous recognition is most often raised as a problem. The paper use a practical example to show how changing perceptions led to a substantial change in research outcomes.

Conclusions

The statistical problem of spontaneous recognition is an unhelpful chimaera encouraging the under-utilisation of valuable data. The problem that should be addressed is one of identity confirmation, which is management issue. A change in both language and attitudes, a focus on the exact nature of the problem being raised, and the use of evidence can generate substantial dividends for both data providers and users. It also suggests a productive way forward for dealing with giving access to non-traditional users such as private sector organisations or journalists.