Letters to the Editor

Data sharing: a decade since the publication of the first cohort profile

From Zhiqiang Wang,* Bin Dong, Odewumi Adegbija, Maryam Sina, Luke Arnold, Tania Pan and Jie Hu

School of Medicine, University of Queensland, Herston, QLD, Australia

*Corresponding author. School of Medicine, University of Queensland, Room 817, Health Sciences Building, RBWH, Herston, QLD, Australia. E-mail: z.wang@uq.edu.au

It has been a decade since the section ‘Cohort Profile’ in the *International Journal of Epidemiology* was launched in December 2004, aiming to foster collaboration through data sharing.¹ To understand the current data sharing practice related to those cohorts, we reviewed 791 subsequent articles which had cited nine early published cohort profiles in 2004/05,²–¹⁰ each cohort containing data from a single nation. By scrutinizing the methods sections of those articles, we identified 606 articles which had used the original data from the profiled cohorts. Data sharing was indicated if the first/corresponding author’s affiliations of a subsequent article were different from those of the original data holders at the organizational level. As shown in Figure 1, the majority of their subsequent publications (73%, 440/606) were produced by the original teams, and about 27% (166/606) were data-sharing articles led by researchers outside the original data holder’s organizations. Over the past decade, data sharing occurred in eight of the nine profiled cohorts. Although data sharing has occurred, the numbers of data-sharing articles varied considerably among those cohorts, ranging from 0 to 49 articles over a decade. Fewer than half (four of nine) of the cohort datasets were shared internationally and total international data-sharing articles constituted about 11% (67/606) of all subsequent articles.

We also found that the original teams which participated in international data sharing produced a greater number of articles by their own teams (non-data sharing articles) than those without international data sharing. On
average, the original teams who had shared their data internationally produced 94 non-data sharing articles per cohort, much more than those of cohorts without international data sharing (13 articles per cohort). Although we could not distinguish between cause and effect, the coexistence of higher levels of data sharing and higher original team’s productivity implies that data sharing benefits both data holders and re-users and is ultimately more valuable for science. The perceived fear of being ‘scooped’ by others using their data before they can is perhaps unfounded.

Original data holders co-authored most subsequent data-sharing articles (85%). However, there was a huge variation among those cohorts. For four of the nine original cohorts, data holders were listed as co-authors in all subsequent data-sharing articles (100%, 84/84). For one of those cohorts, original data holders only co-authored 59% articles and the contribution of original data to the article was only acknowledged in the acknowledgement and methods sections in the remaining 41% articles. We feel that such a format of acknowledgement may undervalue the original data holders’ contribution, considering their efforts to maintain a cohort and the potential research opportunities the contribution can create. Therefore, current recognitions of original data holders’ contributions are still inconsistent and further efforts are needed to develop consensus approaches to crediting original data holders.

Funding
This work was supported by the National Health and Medical Research Council (NHMRC) of Australia (APP1042343 to Z.W.).

Conflict of interest: None declared.

References