Unique identifiers needed to make national datasets fit for public health purposes: the example of subsequent teenage pregnancy in England and Wales

Words 1470

Introduction

The substantial decline in teenage pregnancies in England and Wales over the past few decades has been heralded as a major public health success story. Teenage pregnancies are now at an all-time low, with the conception rate for women aged under 20 nearly halving since 1998 (from 65.1 conceptions per thousand women aged 15-19 in 1998 to 37.9 in 2014) (1,2). However, this conceals differences across different subgroups of young women. While it is recognised that some women will have more than one pregnancy as teenagers, complete and accurate information regarding the extent of subsequent teenage pregnancy remains limited.

Available administrative data suggest that around one quarter of teenage pregnancies in England and Wales are subsequent pregnancies. Using record level data from the Department of Health, McDaid and colleagues (3) showed that 22.9% of women aged under 20 who presented for an abortion in 2013 had been pregnant before. The figures also show that between 1992-2013 the number of young women presenting for an abortion who had been pregnant previously had risen by 30.0% (from 17.2% in 1992 to 22.9% in 2013) (3).

Most of this increase occurred prior to 2004 and since then the proportion of teenagers who have had more than one pregnancy has remained stable. The accuracy of these data may be compromised as they will depend on a woman's willingness to disclose her previous pregnancies and provider records, which may not include details of pregnancies managed by a different hospital or clinic (4).

Changes made in May 2012 to the Population Statistics Act 1938 now require information on the number of previous live-born children be collected for all mothers at birth registration rather than just from married women(5). Figures for 2015 show that 21.4% of

1

young women aged under 20 registering a birth had previously had one or more live-born children(6). However, some young women may include their current birth which will inflate the number of previous births. Moreover, neither of the birth or abortion datasets capture all possible previous pregnancy outcomes patterns.

Pregnancy related data are recorded in various national patient-based datasets. For example, Hospital Episode Statistics contain information about the number of abortions that are carried out in a secondary care setting but in England and Wales only a minority of abortions take place in these environments. In 2015, two thirds (66%) of abortions to young women under 20 were carried out in an independent clinic (7). Primary care datasets (e.g. the General Practice Research Database (GPRD), the Health Improvement Network (THIN) and QResearch) rely on the Read code classification system for patient-related data. Identifying a woman's obstetric history can be challenging, as many codes relating to or suggestive of pregnancy can be used on multiple occasions during one pregnancy. Algorithms have been developed to detect pregnancy history (8,9,10). With all these methods there are difficulties identifying date of conception and duration of gestation for outcomes other than live births. Furthermore, if a woman exercises her right to confidentiality when seeking an abortion, abortion providers are unable to inform the woman's general practitioner (GP) so this information may not appear on her GP medical records (8).

Data derived from statutory abortion notification and birth registration forms are considered to be the most complete data on abortions and births in the UK. Birth registrations are collected under the Births and Deaths Registration Act (1953) and abortion notifications are received under the Abortion Act (1967), amended by the Human Fertilisation and Embryology Act (1990). A project was established to link together ONS birth and Department of Health abortion data to provide an improved, comprehensive epidemiological picture of the proportion of teenagers who have more than one pregnancy, their interpregnancy intervals and any associations with area deprivation.

A unique linked dataset using abortion and birth data

Only two personal identifiers appear on both abortion notification and birth registration data; the young woman's *date of birth* and *postcode* at pregnancy outcome. The abortion notification form (HSA4) was changed in 2001 to require the practitioner completing the termination of pregnancy to include patient name <u>or</u> reference number (patient's hospital <u>or</u> clinic number <u>or</u> NHS number). However, the use of patient reference number rather than name is encouraged and independent clinics are far less likely to use NHS number. The Department of Health use name and/or reference number to check forms in order to monitor the Abortion Act, but neither are stored in electronic datasets. In contrast, all birth registrations include mother's name and are linked by the ONS with birth notification data held by the Personal Demographic Service which includes NHS number.

Working with the Department of Health abortion statistics team, and using data from 2004-2013, attempts were made to link abortion records for England and Wales to birth records for young women aged under 20, creating a unique new dataset. Matching was carried out by the Department of Health using the young woman's date of birth and postcode at pregnancy outcome. Where date of birth and postcode was the same for more than one record it was assumed that the records related to the same woman and the pregnancies were linked. Of the 781,495 pregnancy outcomes in the combined 10-year dataset, only 59,672 or 7.6% young women with more than one pregnancy were identified. This is significantly lower than that reported in available published data.

Like with any exact matching approach, this failed to match records if the young woman's date of birth or postcode was recorded incorrectly, or missing; if they had changed name or address between conceptions; which becomes increasingly likely as the young women get older and leave home to go to college, university and work, or move in with a partner. Furthermore, there were unexpected irregularities in the data variation from year-to-year. For example, in eight years, the number of young women giving birth and identified as having previous pregnancy ranged from 1,962-2,777; but for two years the totals were 35 and 18; and no apparent reason for this has yet been determined.

In order to improve the matching, further 'fuzzy' matching (i.e. matching a partly different postcode to a complete date of birth or vice versa) was considered; however the

3

Department of Health advised discontinuing with this process following the poor exact match rate.

Linking the most complete and accurate national datasets for abortion and birth data had seemed like the most robust approach to identifying patterns of subsequent teenage pregnancy. However this failed because of the inability to reliably match individuals across the two datasets. Without a common unique identifier, such as the NHS number, reliance has to be on a limited set of sociodemographic variables; a method now identified as unsatisfactory and producing inaccurate results.

Implications

The purpose of knowing accurately the levels of subsequent teenage pregnancy is not to provoke further stigma or reduce these young women to statistics, rather it is required to inform public health policies and to help providers of abortion, maternity and sexual health services to better plan and deliver their work. Available data show that while there is a downward trend in overall teenage pregnancies, a downward trend in the risk of subsequent pregnancies cannot be confirmed. The data linkage exercise, though ultimately futile, has identified that no method is yet available to routinely identify the numbers of teenagers who have more than one pregnancy in England and Wales. This is because the information collected is restricted to that required for basic administrative purposes. Without a common unique personal identifier on ONS birth data and Department of Health abortion data it is not possible to accurately link the two datasets. The lack of a common unique identifier, such as NHS number, on abortion data also limits opportunities to link these data with other public health data. This has been recognised in Scotland where abortion notifications are undergoing a one-off retrospective CHI seeding (Community Health Index Number; CHI is the equivalent of NHS Number in England) going back to 1992, which will enable the number of pregnancies and their outcomes to be accurately identified for individuals for research and monitoring purposes. While the context in Scotland is different to England and Wales, in so far as the majority of abortions take place in NHS hospitals making this easier to put into practice, this does set precedence for the interpretation of data protection legislation. In Denmark, all induced abortions from its Hospital Discharge Register (LPR; Landspatientregister), include women's identification

4

numbers, while Finland has a Register on Induced Abortions and Sterilisations which makes it possible to combine with other data sources.

We therefore advocate a change in routine data collection to include NHS number on all abortion notification forms and stored on datasets so that this can be used, in combination with other personal identifiers, to make this data more useful for detecting trends and research purposes.

References

1. Office of National Statistics (ONS). Conceptions in England and Wales. London: ONS. Avaliable at: http://webarchive.nationalarchives.gov.uk/20160105160709/http:// ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-39646, 2010 (last accessed 02/20/2017).

2. Office of National Statistics (ONS). Conceptions in England and Wales, 2014. London: ONS. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandma rriages/conceptionandfertilityrates/bulletins/conceptionstatistics/2014; 2016 (last accessed 02/20/2017).

3. McDaid LA, Collier J, Platt MJ. Previous pregnancies among young women having an abortion in England and Wales. J Adolescent Health. 2015; 57:387-92.

4. Gbolade BA. Teenage pregnancy rates and the age and sex of general practitioners: Record linkage analysis could have been used. BMJ. 2000; 321:381.

5. Office of National Statistics (ONS). How changes to the Population Statistics Act will affect birth statistics. London: ONS. Available at: http://webarchive.nationalarchives.gov.uk/ 20160105160709/http://www.ons.gov.uk/ons/guide-method/user-guidance/health-and-life-events/how-changes-to-population-statistics-act-will-affect-birth-statistics.pdf; 2012 (last accessed 02/20/2017).

6. Office of National Statistics (ONS). Births by Parents' Characteristics 2015, dataset. London: ONS. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/ birthsdeathsandmarriages/livebirths/datasets/birthsbyparentscharacteristics; 2016 (last accessed 02/20/2017).

7. Department of Health. Abortion statistics, England and Wales: 2015, statistical dataset. London: Department of Health. Available from: https://www.gov.uk/government/statistical-data-sets/abortion-statistics-england-and-wales-2015; 2016 (last accessed 02/20/2017).

8. Devine S, West S, Andrews E, Tennis P, Hammad TA, Eaton S, et al. The identification of pregnancies within the general practice research database. Pharmacoepidem Dr S. 2010; 19:45-50.

9. Hardy JR, Holford TR, Hall GC, Bracken MB. Strategies for identifying pregnancies in the automated medical records of the General Practice Research Database. Pharmacoepidem Dr S. 2004; 13:749-59.

10. Manson JM, McFarland B, Weiss S. Use of an automated database to evaluate markers for early detection of pregnancy. Am J Epidemiol. 2001; 154:180-7.