Supplement 3: Estimation of Prescription Lengths

In 344,720 pregnancies where depressive symptoms or prescription of an antidepressant were recorded, we observed 6,821,177 prescribing events that had occurred up to a year before or during pregnancy. To estimate the periods during which antidepressants had been prescribed, we required information on the date of prescription, the dose unit, the prescribed quantity, and the recommended daily dose. Because prescribers can record information on patients' medical records in various ways (e.g. using a free text description rather than numerical fields) there were missing values within the allocated fields for these variables. As such, there were 5,821,411 records (85.3%) where the dose unit was not recorded; 40,269 records (0.6%) where the prescribed quantity was not recorded or recorded as zero; and 897,234 records (13.2%) where the recommended daily dose was not recorded or recorded as zero. Our first aim was, therefore, to collate this information from other sources within the therapy record.

Specifically:

- We derived missing values on dose unit from the product name or the pack type description
- We derived missing values on prescribed quantity from free text descriptions, or from the pack type description and number of packs prescribed
- We derived missing values on the recommended daily dose from the number of dose units to be taken on a single occasion together with daily frequency, or from free text descriptions

Where plausible values for dose unit, daily dose, or prescribed quantity could not be collated from other sources within the therapy record, they were left as missing. Following the above procedure, there remained 32,344 records with missing values on dose unit (0.5%); 40,127 records with missing values on prescribed quantity (0.6%); and 870,638 records with missing values on recommended daily dose (12.8%).

At this point we derived a preliminary prescription length variable by dividing the prescribed quantity by the recommended daily dose and investigated the resulting prescription length distribution to identify further potential recording errors. Records for which the derived prescription length was unusually short, unusually long or otherwise implausible were flagged as potential recording errors. For these flagged records, we cross-referenced all available information within the therapy record specifically to examine the following conditions:

- Records where the prescribed quantity was equal to, or smaller than the daily dose
- Prescriptions where the derived prescription length suggested a duration of less than one week
- Prescription records with a non-integer value for the derived prescription length
- Prescriptions records with length of more than 90 days

For each of these conditions, we devised an algorithm to identify and amend likely recording errors by cross-referencing with other information within the therapy record. This procedure was carried out in an iterative fashion where: (i) a prescription length variable was derived; (ii) potential recording errors were identified and scrutinised; (iii) our algorithm was adjusted to amend these recording errors; (iii) a new prescription length variable was derived; (iv) potential recording errors were identified; and so on. We carried out this procedure until no further potential recording errors were identified. Where implausible values could not be amended using within information from the therapy record, they were set to missing. Following this procedure, there were 101,164 records (1.5%) with likely errors in prescribed quantity, and 871,179 records (12.8%) with likely errors in recommended daily dose which could not be amended using other information from the therapy record.

For these remaining records with missing values on prescribed quantity or recommended daily dose we imputed values using the following rule hierarchy:

1) Impute the modal value of records pertaining to the same patient, the same product code and dose unit, and the same recommended daily dose or prescribed quantity.
2) Impute the modal value of records pertaining to other patients, but the same product code and dose unit, and the same recommended daily dose or prescribed quantity.
3) Impute with the modal value of records pertaining to other patients, but the same product code and dose unit.
4) Impute with the modal value of records pertaining to other patients, but with the same product name and dose unit.
5) Impute with the modal value of records pertaining to other patients, but with the same drug substance name and dose-unit.

Using this approach, we imputed 838,835 missing values on recommended daily dose (n=355,224 with rule 1; n=478,312 with rule 2; n=5,274 with rule 3; none with rule 4; and n=25 with rule 5) and 68,820 missing values on prescribed quantity (n=49,324 with rule 1; n=19,365 with rule 2; n=92 with rule 3; none with rule 4; and n=39 with rule 5). The remaining 32,344 records (0.5%) with missing values on dose unit, prescribed quantity and recommended daily dose were not used in the estimation of prescription periods.

We provide a visual comparison of the prescription length distributions based on records with complete versus imputed data in the figure below.