The method used, and software, was adapted from that designed by Richard Peto, Richard Doll and Martin Vessey in their Oxford-based cohort studies of women using different contraceptive methods.\textsuperscript{1–3} In the contraceptive study (which ran from the 1970s to 2013, with the methods extensively tested and used), the investigators compared women in different contraceptive user groups in strata of 4 age groups, 2 parity groups, 3 social class groups, and 3 smoking groups. With 17,032 subjects, and 72 strata (4x2x3x3), this gave an average of 236 people per stratum. In the present HIV/MS study we had 16,200 strata (18 ages, 2 sexes, 10 calendar years, 9 regions, 5 quintiles of deprivation) which, with 5,319,703 subjects, which gives an average of 328 people in each stratum (a greater number than in the original Doll/Peto study). Analyses were run using SAS (release 9.2, SAS Institute Inc., Cary, NC, USA).

The key numbers are those in the combined HIV and reference cohort (dominated numerically by the reference cohort). This is the ‘standard population’. The stratum-specific rates in the standard population were applied to each equivalent stratum in, first, the HIV cohort and, second, the reference cohort. Taking the largest HIV age stratum (aged 35–39), subdivided into an additional 900 strata (2 sexes x 10 years x 9 regions x 5 deprivation quintiles), the analysis was based on 449,754 people (average 500 people per stratum); taking the smallest (aged 80–84), the analysis was based on 282,955 people (average 314 people per stratum). Some strata in the standard population will have no people in the stratum: the stratum contributes nothing to either the expected number or (by definition) to the observed number. Rates in strata in the standard cohort that include numbers will give a tiny stratum-specific rate. We calculate this to a very large number of significant digits – up to 16, when 16 is arithmetically possible - to ensure that the sum of the rates is very exact. Summing across all strata gives an expected number (e.g. 18.3 cases of MS in the HIV cohort in the main calculation in the manuscript). The observed number – 7 – is simply the overall count in the HIV cohort and is unmodified by any of the stratification and standardising procedures or calculations.

We always run each individual condition in the reference cohort against all others combined. In the HIV/MS analysis, for example, the appendicitis sub-cohort, run for MS as an outcome, compared with all other conditions in the reference cohort combined, has a rate ratio (RR) for MS of 1.01; the cohort of people with internal derangement of the knee has a RR for MS of 1.02; bunion is 0.97; contraceptive management is 0.92; gall bladder disease is 1.08; inguinal hernia is 0.91. In other words, the stratified analyses show that there is neither any important increase nor decrease in the risk of MS in these sub-cohorts. This indicates that the method works in cohorts where there is no reason to expect an increased or decreased risk of MS. It gives rates that approximate to 1, where 1 is expected.

We considered the possibility that there could be a distorting effect in using strata with relatively small numbers of cases in the HIV cohort. We therefore ran the analyses again excluding people aged over 70 (table 1 in the manuscript). In this “sensitivity” analysis the number of observed cases of MS in people less than 70 years old was 6, the number expected was 17.7, and the rate ratio was 0.34 (95% CI 0.12 to 0.74).
