

Methods S1. Supplementary method for calculating the sampling rate

Initially, we intended to request the Korea Disease Control and Prevention Agency to supply a random sample of 1% from tertiary general hospitals, 5% from general hospitals, and 10% from other hospitals for cases without the International Classification of Diseases (ICD) codes. However, due to a miscommunication, the provided sampling was based on the number of cases with the ICD codes, which led us to adjust our methodology for calculating the sampling rates.

The total sampling rates for cases categorized as algorithm-positive with the ICD codes and algorithm-negative with the ICD codes were calculated as originally planned. However, the total sampling rate for algorithm-negative cases without the ICD codes required modification.

Since the exact number of cases in this group was unknown, we approximated the number of the control group as algorithm-negative cases without the ICD codes for the calculation of the sampling rate. This estimation was made using the proportion (k value) of stroke and acute myocardial infarction patients, derived from the National Health Insurance Service-National Sample Cohort data in 2013.

With the estimated number and the k value, we were able to calculate the total sampling rate (Figure S1) for each stratum, taking into account the algorithm-negative cases without the ICD codes. This adjustment was necessary due to the lack of specific data for algorithm-negative cases without the ICD codes from the Korea Disease Control and Prevention Agency.