**SUPPLEMENTAL APPENDIX**

**Table 5.** Newcastle-Ottawa Scale (NOS) ratings for the 14 included studies.

| **Reference** | **Selection of Patients:**  representativeness of the exposed cohort, selection of the non-exposed cohort, ascertainment of exposure, and demonstration that outcome of interest was not present at the start of the study | **Comparability of Study Groups:**  comparability of cohorts on the basis of the design or analysis | **Determination of Outcomes of Interest:**  assessment of outcome, adequate follow-up time for outcomes to occur, adequacy of cohort follow up | **Overall NOS Rating** |
| --- | --- | --- | --- | --- |
| Chen Acta Neurol Scand 2014 [18] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Chen Circ J 2016 [19] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Chinwong, D. Ther Clin Risk Manag 2015 [20] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Danese BMJ Open 2016 [21] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Huang Am Health Drug Benefits 2016 [22] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Jena Am J Manag Care 2016 [23] | \*\*\*\* | \*\* | \*\* | 8 |
| Leibowitz JAMA Intern Med 2016 [24] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Lin Int J Cardiol 2017 [25] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Ostergaard Neuroepidemiology 2014 [26] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Sheng Eur J Clin Pharmacol 2012 [27] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Sheng QJM 2012 [28] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Sicras-Mainar J Clin Pharm Ther 2012 [29] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Smith Am J Med 2015 [30] | \*\*\*\* | \*\* | \*\*\* | 9 |
| Toth J Med Econ 2017 [31] | \*\*\*\* | \*\* | \*\*\* | 9 |

The NOS assessment evaluates studies on the following three categories: selection of participants, the comparability of study groups, and the determination of outcomes of interest. Each category is evaluated based on category-specific criteria. The selection category is evaluated based on the following criteria: representativeness of the exposed cohort, selection of the non-exposed cohort, ascertainment of exposure, and demonstration that outcome of interest was not present at the start of the study. Comparability is evaluated based on the comparability of cohorts on the basis of the design or analysis. Evaluation criteria for the outcome category includes assessment of outcome, adequate follow-up time for outcomes to occur, and adequacy of cohort follow up. The NOS uses a star rating system to rate the criteria for each of the three categories. Except for comparability, each criterion receives a single star if appropriate methods have been reported. Comparability may receive two stars: one if the most important confounders were adjusted for the analysis, and two if any other adjustments were made. The maximum number of nine stars (selection [4 stars], comparability [2 stars], outcome [3 stars]) indicates a high quality study. A NOS score of ‘0’ stars indicates a low quality study.