

## Appendix 2. Description of Study Cohorts Used in Plasmode Simulations

## **Appendix 2: Description of Study Cohorts Used in Plasmode Simulations**

1. ***Cohort 1 (NSAID dataset)***: Data consists of 49,653 Medicare beneficiaries ages 65 and older who were enrolled in the Pharmaceutical Assistance Contract for the Elderly (PACE) program provided by the state of Pennsylvania. The study cohort included individuals who initiated either a Cox-2 inhibitor or a nonselective NSAID between 1999-2002. Initiation of NSAID or Cox-2 inhibitor was defined as filling at least one prescription for an NSAID, or Cox-2 inhibitor, between January 1, 1999, and December 31, 2002, without any prescription of an NSAID, or Cox-2 inhibitor, during the 18 months prior to the index date (new-user design). Individuals were followed for 180 days after initiation of therapy with the study outcome being gastrointestinal complications. The data consists of 9,470 unique codes within eight data dimensions (prescription drugs, ambulatory diagnoses, hospital diagnoses, nursing home diagnoses, ambulatory procedures, hospital procedures, physician diagnoses, and physician procedures). Each code represents the frequency of times the code appeared in the pre-specified washout period (18 months prior to the index date). These codes were used to create HDPS generated variables as described in the manuscript.
2. ***Cohort 2 (NOAC dataset)***: Data were collected by United Healthcare and consists of 18,447 individuals who initiated either a novel oral anticoagulant (NOAC) or warfarin between October 2009 and December 2012 after a 1 year washout period with no prescription for any oral anticoagulant (new-user design). Data were collected to compare the effect of warfarin versus NOAC on a number of outcomes, including a combined outcome of all-cause mortality, ischemic stroke, and hemorrhagic stroke. Data consisted of 23,531 unique claims codes within 5 data dimensions (inpatient procedures, outpatient procedures, inpatient diagnoses, outpatient diagnoses, and prescription drugs). Each code represents the frequency of times the code appeared in the pre-specified washout period (12 months prior to the index date). These codes were used to create HDPS generated variables as described in the manuscript.
3. ***Cohort 3 (Vytorin dataset)***: Data consists of a population of elderly adults (>65 years) who were enrolled in Pennsylvania's Pharmaceutical Assistance Contract for the Elderly (PACE) program provided by the state of Pennsylvania. The dataset consists of 36,122 individuals who initiated either a statin or glaucoma medication after no prescription of either a statin or glaucoma medication within the 12 months prior to the index date (new-user design). Individuals were followed to evaluate the protective effect of initiating statin therapy on all-cause mortality. The data consists of 9,470 unique codes within eight data dimensions (prescription drugs, ambulatory diagnoses, hospital diagnoses, nursing home diagnoses, ambulatory procedures, hospital procedures, physician diagnoses, and physician procedures). Each code represents the frequency of times the code appeared in the pre-specified washout period (12 months prior to the index date). These codes were used to create HDPS generated variables as described in the manuscript.