Appendix A – UnitedHealth Group Clinical Research Database

Standardization of Data Entry and Data Structure

Medical and pharmacy claims data are captured, predominantly electronically, from sites of care seeking third-party reimbursement for both Medicare and commercial plans using the industry standard data collection forms HCFA/CMS-1500 for facility claims, UB04/CMS-1450 for professional services and outpatient claims, and NCPDP for pharmacy claims or their electronic equivalents. Structured data from these standardized forms are coded using the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), National Drug Codes (NDC), Current Procedural Terminology (CPT) codes, and Logical Observation Identifiers Names and Codes (LOINC) codes, and Diagnosis Related Groups (DRG). This nomenclature ensures consistency of data collection across geographic regions, health systems, and payers throughout the United States.

Methods to Control for Errors in Sampling and Data Collection

Claims that do not adhere to the form or coding standards described above are rejected from reimbursement, minimizing the risk that inappropriately structured data are included in the database. Data specific to SARS-CoV-2 and COVID-19 has an additional quality control layer to control for errors in sampling and data collection; this is described below in the section on Quality Control.

Quality Control

A COVID-19 data source-specific layer of quality control is also present, given the rapidly evolving situation. SARS-CoV-2 lab tests included in the database exclude custom local codes or

codes that are not present in the LOINC organization's guidance for mapping SARS-CoV-2 and COVID-19 related LOINC terms. Test information provided via the LOINC code compliments the test type (antibody, RT-PCR, etc.) as well as the result value (detected, not detected, not given/cancelled). Suspected COVID-19 inpatient cases are manually reviewed daily by clinical staff via clinical notes to determine an individual's COVID-19 status. Each case is then manually flagged as either negative, confirmed, presumed positive, or needs clinical review. If a case is confirmed, it is not reviewed again. If a case is listed as negative or unknown, it is periodically reviewed for changes in the record. All others are reviewed and updated daily.

Data Sharing

The data are proprietary and are not available for public use but, under certain conditions, may be made available to editors and their approved auditors under a data use agreement to confirm the findings of the current study. Data that are made available for research through the UHG Clinical Discovery Portal use the de-identified primary key as the link across data tables. All protected information has been removed, ensuring any research performed is limited to retrospective analysis of de-identified data and accessed in accordance with Health Insurance Portability and Accountability Act regulations.