

## 1.1 Description of Data Fields

Below we describe the fields of the three data sets provided in the Heritage Health Prize. For information regarding the generalization of data elements, please see the following sections of this document. Note that pseudonyms are used for all identification variables (MemberID, ProviderID, Vendor, and PCP), based on the generation of sets of random numbers that were then assigned randomly. Ordering the data using these pseudonyms will not order the data in any meaningful way. The rows of each data set were randomly shuffled to further ensure that the order of the pseudonyms is not mistakenly taken to represent the order of the original data.

All claims that appear within the same DSFS interval may be in any order. It is only possible to look at disease progression on a monthly basis.

In the original data close to 96% of the claims had no LengthOfStay value. Therefore, that was how the data looked at the outset and the missing values were not due to suppression. We have added a binary field (SupLOS) to make that clear.

### Members Data

MemberID	Member pseudonym.
AgeAtFirstClaim	Age in years at the time of the first claim's date of service computed from the date of birth; Generalized into ten year age intervals.
Sex	Biological sex of member: M = Male; F=Female.

### Claims (Level 2) Data

MemberID	Member pseudonym.
ProviderID	Provider pseudonym.
Vendor	Vendor pseudonym.
PCP	Primary care physician pseudonym.
Year	Year in which the claim was made: Y1; Y2; Y3.
Specialty	Generalized specialty.
PlaceSvc	Generalized place of service.
PayDelay	Number of days delay between the date of service (the date the actual procedure was performed or service provided) and date of payment. Values above 161 days (the 95% percentile) are top-coded as "162+".
LengthOfStay	Length of stay (discharge date – admission date + 1), generalized to: days up to six days; (1-2] weeks; (2-4] weeks; (4-8] weeks; (8-12 weeks]; (12-26] weeks; more than 26 weeks (26+ weeks).
DSFS	Days since first claim, computed from the first claim for that member for each year, generalized to: [0-1] month, (1-2] months, (2-3] months, (3-4] months, (4-5] months, (5-6] months, (6-7] months, (7-8] months, (8-9] months, (9-10] months, (10-11] months, (11-12] months.

PrimaryConditionGroup	Broad diagnostic categories, based on the relative similarity of diseases and mortality rates, that generalize the primary diagnosis codes (ICD-9-CM) [2].
CharlsonIndex	A measure of the affect diseases have on overall illness, grouped by significance, that generalizes additional diagnoses. Scores greater than zero are carried forward (for up to a year) in subsequent claims with a co-morbidity score of zero [1,4].
ProcedureGroup	Broad categories of procedures, grouped according to the hierarchical structure defined by the Current Procedural Terminology (CPT) [3].
SupLOS	Indicates if the NULL value for the LengthOfStay variable is due to suppression done during the de-identification process. A value of 1 indicates that suppression was applied.

#### Drug Count Data

MemberID	Member pseudonym.
Year	Year in which the drug prescription was filled: Y1; Y2; Y3.
DSFS	Days since first service (or claim), computed from the first claim for that member for each year, generalized to: [0-1] month, (1-2] months, (2-3] months, (3-4] months, (4-5] months, (5-6] months, (6-7] months, (7-8] months, (8-9] months, (9-10] months, (10-11] months, (11-12] months.
DrugCount	Count of unique prescription drugs filled by DSFS. No count is provided if prescriptions were filled before DSFS zero. Values above 6, the 95% percentile after excluding counts of zero, are top-coded as "7+".

#### Lab Count Data

MemberID	Member pseudonym.
Year	Year in which the drug prescription was filled: Y1; Y2; Y3.
DSFS	Days since first service (or claim), computed from the first claim for that member for each year, generalized to: [0-1] month, (1-2] months, (2-3] months, (3-4] months, (4-5] months, (5-6] months, (6-7] months, (7-8] months, (8-9] months, (9-10] months, (10-11] months, (11-12] months.
LabCount	Count of unique laboratory and pathology tests by DSFS. Values above 9, the 95% percentile after excluding counts of zero, are top-coded as "10+".

#### Outcome Data

MemberID	Member pseudonym.
DaysInHospital_Y2	Days in hospital, the main outcome, for members with claims in Y1. Values above 14 days (the 99% percentile) are top-coded as "15+".
DaysInHospital_Y3	Days in hospital, the main outcome, for members with claims in Y2. Values above 14 days (the 99% percentile) are top-coded as "15+".
ClaimedTruncated	Members with truncated claims in the year prior to the main outcome are assigned a value of 1, and 0 otherwise.

## 1.2 Grouping CPT Codes into Procedure Groups

Procedure groups refer to broad categories of procedures, grouped according to the hierarchical structure defined by the Current Procedural Terminology [3].

<b>CPT GROUP</b>	<b>DESCRIPTION &amp; INCLUDED CPT CODES</b>
<u><b>ANESTHESIA</b></u>  ANES	Head, neck, thorax, intrathoracic, spine and spinal cord, upper and lower abdomen, perineum, pelvis, upper and lower leg, knee and popliteal area, shoulder and axilla, upper arm and elbow, forearm, wrist and hand, radiological procedures, burn excisions or debridement, obstetric, and other.  00100-01999
<u><b>EVALUATION AND MANAGEMENT</b></u>  EM	Office or other outpatient services, hospital observation services, hospital inpatient services, consultations, emergency department services, critical care services, continuing intensive care services, nursing facility services, domiciliary, rest home, or custodial care services, home services, prolonged services, case management services, care plan oversight services, preventive medicine services, special evaluation and management services, and other manual and management services.  99201-99499
<u><b>MEDICINE</b></u>  MED	All other medicine including drug administration, vaccines, toxoids, hydration, therapeutic, prophylactic, and diagnostic injections and infusions, psychiatry, dialysis, ophthalmology, contact lens services, spectacle services, medical tests and measurements, analysis, assessment, intervention, evaluative and therapeutic services, diagnostic studies, drug administration, physical medicine and rehabilitation, education and training for patient self-management, special services, procedures and reports, moderate sedation, and home health procedures and services.  90281-99199, 99500-99602
<u><b>PATHOLOGY AND LABORATORY</b></u>  PL	Organ or disease panels, drug testing, therapeutic drug assays, evocative and suppression testing, consultations, urinalysis, chemistry, molecular diagnostics, infectious agent: detection of antibodies, microbiology infectious agent detection, anatomic

	pathology, cytopathology, cytogenetic studies, and surgical pathology.  80048-89356
<b><u>RADIOLOGY</u></b>  RAD	Diagnostic radiology, diagnostic ultrasound, radiation oncology, and nuclear medicine.  70010-79999
<b><u>SURGERY-AUDITORY SYSTEM</u></b>  SAS	External ear, middle ear, inner ear, and temporal bone, middle fossa approach.  69000-69979
<b><u>SURGERY-CARDIOVASCULAR SYSTEM</u></b>  SCS	Heart and pericardium, and arteries and veins.  33010-37799
<b><u>SURGERY-DIGESTIVE SYSTEM</u></b>  SDS	Lips, vestibule of mouth, tongue and floor of mouth, dentoalveolar structures, palate and uvula, salivary gland and ducts, pharynx, adenoids, and tonsils, esophagus, stomach, intestines, meckel's diverticulum and the mesentery, rectum, anus, liver, biliary tract, pancreas, abdomen, and peritoneum, and omentum.  40490-49999
<b><u>SURGERY-EYE AND OCULAR ADNEXA</u></b>  SEOA	Eyeball, anterior segment, posterior segment, ocular adnexa, and conjunctiva.  65091-68899
<b><u>SURGERY-GENITAL SYSTEM</u></b>  SGS	Male: penis, testis, epididymis, tunica vaginalis, scrotum, spermatic cord, seminal vesicles, and prostate; Female: vulva, perineum and introitus, vagina, cervix uteri, corpus uteri, oviduct and ovary, ovary, and in vitro fertilization.  54000-55899, 56405-58999
<b><u>SURGERY-INTEGUMENTARY SYSTEM</u></b>  SIS	Skin, subcutaneous and accessory structures, nails, pilonidal cyst, introduction, repair, destruction, and breast.  10040-19499

<u><b>SURGERY-MATERNITY CARE AND DELIVERY</b></u>  SMCD	Maternity care and delivery.  59000-59899
<u><b>SURGERY-MUSCULOSKELETAL SYSTEM</b></u>  SMS	General, head, neck and thorax, back and flank, spine, humerus and elbow, forearm and wrist, hand and fingers, pelvis and hip joint, femur and knee joint, foot and toes, application of casts and strapping, and endoscopy and arthroscopy.  20000-29999
<u><b>SURGERY-NERVOUS SYSTEM</b></u>  SNS	Skull, meninges, and brain, spine and spinal cord, and extracranial nerves, peripheral nerves, and autonomic nervous system.  61000-64999
<u><b>SURGERY-OTHER</b></u>  SO	Surgery-endocrine system: thyroid gland, and parathyroid, thymus, adrenal glands, pancreas, and carotid body; Surgery-hemic and lymphatic systems: spleen, general, lymph nodes and lymphatic channels; Surgery-mediastinum and diaphragm; Surgery-operating microscope.  60000-60699, 38100-38999, 39000-39599, 69990-69990
<u><b>SURGERY-RESPIRATORY SYSTEM</b></u>  SRS	Nose, accessory sinuses, larynx, trachea and bronchi, and lungs and pleura.  30000-32999
<u><b>SURGERY-URINARY SYSTEM</b></u>  SUS	Kidney, ureter, bladder, and urethra.  50010-53899

### 1.3 Grouping ICD-9-CM Codes into Primary Condition Groups

Primary conditions refer to broad diagnostic categories, which are based on the relative similarity of diseases and mortality rates [2].

<b>PRIMARY CONDITION GROUP</b>	<b>DESCRIPTION &amp; INCLUDED ICD CODES</b>
<b><u>ACUTE MYOCARDIAL INFARCTION</u></b>  AMI	Myocardial infarction.  410-414
<b><u>ACUTE RENAL FAILURE</u></b>  RENAL1	Acute renal failure, nephrotic syndrome, and related conditions.  580, 581, 584
<b><u>ACUTE RESPIRATORY</u></b>  RESPR4	Acute respiratory infections and miscellaneous respiratory diseases.  460-478, 786
<b><u>ALL OTHER INFECTIONS</u></b>  INFEC4	All other infections, joint infections and muscle infections, with the exception of hepatitis; unspecified fever.  001-139, multiple others, incl. joint infections & muscle infections (711 & 728); 780.6 (fever)
<b><u>ALL OTHER TRAUMA</u></b>  TRAUMA	Traumatic injuries not included elsewhere, including head injuries without intracranial or subdural bleeds.  800-804, 840-848, 850-854, 860-904, most of 905-959
<b><u>APPENDICITIS</u></b>  APPCHOL	Appendicitis, hernias, cholecystitis, and cholangitis.  540-543, 550-553, 574-576
<b><u>ARTHROPATHIES</u></b>  ARTHSPIN	Arthropathies and spine disorders (but no infections or autoimmune conditions).  712, 715-729, most 731-739 (except for 733.1x, pathologic fracture)
<b><u>ATHEROSCLEROSIS AND PERIPHERAL VASCULAR DISEASE</u></b>  HEART4	Atherosclerosis (including that affecting precerebral arteries) and other forms of peripheral vascular disease.  429.2, 433, 440-459
<b><u>CANCER A</u></b>  CANCRA	Malignant neoplasms of respiratory tract and intrathoracic organs; leukemias, non-Hodgkin's lymphomas, and other histiocytic malignancies.  160-165, 202-208
<b><u>CANCER B</u></b>	All other malignant neoplasms not in Cancer A or gynecologic ones (including Hodgkin's disease); radiation therapy and

CANCRB	chemotherapy encounters where cancer not specified.  140-159, 170-173, 175, 176, 185-195, 200, 201, V58.0, V58.1, V66.1, V66.2
<b><u>CATASTROPHIC CONDITIONS</u></b>  CATAST	Catastrophic conditions, including dissecting aneurysms, cardiac arrest, respiratory arrest, all forms of shock except septic shock; intracranial and subdural haemorrhages.  Multiple ICD codes
<b><u>CHEST PAIN</u></b>  ROAMI	Chest pain, myocardial infarction not specified.  Misc. 786.5X, V71.7
<b><u>CHROIC OBSTRUCTIVE PULMONARY DISORDER</u></b>  COPD	Chronic obstructive pulmonary disorder and some less common respiratory conditions.  490-496, 500-508, 512, 515, 517-519
<b><u>CHRONIC RENAL FAILURE</u></b>  RENAL2	Chronic renal failure, end-stage renal disease, and kidney transplants.  582, 583, 585-589, 996.81, V42.0xx
<b><u>CONGESTIVE HEART FAILURE</u></b>  CHF	Congestive heart failure and some related illnesses.  Major codes are 425, 428, and miscellaneous (398.91, 402s, 422s, and some 429s, incl. '429')
<b><u>DIABETIC KETOACIDOSIS AND RELATED MATABOLIC</u></b>  METAB1	Diabetic ketoacidosis, with and without coma; hypoglycemic coma; unspecified coma and alteration of consciousness.  Misc. 250s, 251, 780.0x
<b><u>FLUID AND ELECTROLYTE</u></b>  FLaELEC	Typical fluid and electrolyte disorders and dehydration.  275.2 – 276.9
<b><u>FRACTURES AND DISLOCATIONS</u></b>  FXDISLC	All other fractures and dislocations, including pathologic fractures.  733.1xx, 805-807, 809-819, 822-839, misc. 905, 907, 952
<b><u>GASTROINTESTINAL BLEEDING</u></b>  GIBLEED	Gastrointestinal hemorrhage; miscellaneous disorders of stomach and duodenum; diverticulitis; abdominal symptoms, nausea with vomiting; blood in stool.  530-537, 562, 564, 565, 569, 578, 787, 789, 792.1
<b><u>GASTROINTESTINAL, INFLAMMATORY BOWEL DISEASE, AND OBSTRUCTION</u></b>  GIOBSENT	Inflammatory bowel disease and malabsorption; gastrointestinal obstruction; enteritides.  555-558, 560, 568, 579

<b><u>GYNECOLOGIC CANCERS</u></b>  GYNECA	Gynecologic malignancies other than ovarian cancer; female breast cancer.  174, 179-182, 184
<b><u>GYNECOLOGY</u></b>  GYNEC1	Non-malignant, non-infectious gynecologic diseases, including benign neoplasms.  218-221, 256 & multiple miscellaneous codes (including V codes).
<b><u>HIP FRACTURE</u></b>  HIPFX	Hip fracture.  Some 733s, 808, 820, 821, some 905s, 959.6
<b><u>INGESTIONS AND BENIGN TUMORS</u></b>  ODaBNCA	Non-gynecologic benign neoplasms; drug overdoses, drug abuse, adverse drug reactions, and poisonings.  210-217, 222-239, 610, 611; 291, 292, 303-305, 790.3, 796, 960-989, 995.2
<b><u>LIVER DISORDERS</u></b>  LIVERDZ	Liver disorders, including hepatitis.  570-573
<b><u>MISCELLANEOUS # 1</u></b>  MISCL1	Miscellaneous conditions not classified previously.  990-999
<b><u>MISCELLANEOUS # 2</u></b>  MSC2a3	External causes of injury; remaining supplemental classification of factors influencing health status and contact with health services.  Remaining V codes; remaining 790-796; all E codes.
<b><u>MISCELLANEOUS # 3</u></b>  MISCL5	Miscellaneous non-cardiac congenital anomalies; miscellaneous symptoms other than fever; miscellaneous tooth and tongue disorders, miscellaneous diagnoses of pain.  520-529 (tooth & tongue disorders); 740-759, 780 (except for 780.6), 783-785 (if not found elsewhere); 338
<b><u>MISCELLANEOUS CARDIAC</u></b>  MISCHRT	Miscellaneous cardiac conditions and congenital heart disease.  392-405, 745-747
<b><u>NON-MALIGNANT HEMATOLOGIC</u></b>  HEMTOL	Hematologic problems other than malignancies.  273, 280-289, misc 790s, 996.85
<b><u>OTHER CARDIAC CONDITIONS</u></b>  HEART2	Diseases of pulmonary circulation and cardiac dysrhythmias.  415-417, 426, 427, misc. 785s, misc. 996s
<b><u>OTHER METABOLIC</u></b>  METAB3	All other endocrine, metabolic and miscellaneous immune disorders (but not including systemic lupus erythematosus or

	rheumatoid arthritis).  240-255, 257-272, 274-275.1, 277-279, misc. 790s
<b><u>OTHER NEUROLOGICAL</u></b>  NEUMENT	All other neurologic problems and mental disorders (other than drug overdoses); senility.  290-319, 327-344, 346-389, 781, 797, V71.0
<b><u>OTHER RENAL</u></b>  RENAL3	All other renal diseases other than infections.  Miscellaneous 405s, 591-608, misc. other codes
<b><u>OVARIAN AND METASTATIC CANCER</u></b>  CANCRM	Ovarian cancer and metastatic cancer.  183, 196-199
<b><u>PANCREATIC DISORDERS</u></b>  PNCRDZ	Pancreatic disorders.  577
<b><u>PERICARDITIS</u></b>  PERVALV	Pericarditis and valvular heart disease.  391, 423, 424
<b><u>PERINATAL PERIOD</u></b>  PERINTL	All conditions originating in the perinatal period.  760-779
<b><u>PNEUMONIA</u></b>  PNEUM	All forms of pneumonia; empyema; pleurisy; and lung abscess; also includes pulmonary tuberculosis; pulmonary congestion and hypostasis.  480-487; 510; 511; 513; 011, 012.8; 514
<b><u>PREGNANCY</u></b>  PRGNCY	Pregnancy and related conditions, including circumstances related to reproduction and development.  630-677, V22-V28
<b><u>SEIZURES</u></b>  SEIZURE	Seizure disorders.  345, misc. 780.1-780.4
<b><u>SEPSIS</u></b>  SEPSIS	Sepsis, meningitis, septic shock, and major catastrophic infections.  003.1, 003.21, 027.0, 036-038, 040, 320-326, 422.92, 728.86, 785.4, 785.59, 790.7, 995.92, 9993
<b><u>SKIN AND AUTOIMMUNE DISORDERS</u></b>  SKNAUT	SLE, rheumatoid arthritis, skin disorders, & related autoimmune diseases, sialoadenitis  690-710, 713, 714, 782
<b><u>STROKE</u></b>  STROKE	Stroke and post-stroke complications.  434-438, 997.0x

<b><u>URINARY TRACT INFECTIONS</u></b>	Urinary tract infections, not including pregnancy related ones.
UTI	590, 595, 597, 599, 601, 604, misc. 996s

## 1.4 Grouping Place of Service

Places of service, described in [3], were grouped in consultation with hospital physicians.

<b>PLACESVC GROUP</b>	<b>DESCRIPTION &amp; INCLUDED PLACE OF SERVICE CODES</b>
<b><u>AMBULANCE</u></b>	Ambulance: A vehicle specifically designed, equipped, and staffed for lifesaving and transporting the sick or injured; Ambulatory surgical center: a freestanding facility, other than a physician's office, where surgical and diagnostic services are provided on an ambulatory basis.
<b><u>HOME</u></b>	Location, other than a hospital or other facility, where the patient receives care in a private residence.
<b><u>INPATIENT HOSPITAL</u></b>	A facility, other than psychiatric, that primarily provides diagnostic, therapeutic, and rehabilitation services by physicians for admitted patients.
<b><u>INDEPENDENT LAB</u></b>	A laboratory certified to perform diagnostic or clinical tests independent of an institution or a physician's office.
<b><u>OFFICE</u></b>	Location where the health professional routinely provides health examinations, diagnosis, and treatment of illness or injury on an ambulatory basis.
<b><u>OUTPATIENT HOSPITAL</u></b>	A portion of a hospital that provides diagnostic, therapeutic (both surgical and nonsurgical), and rehabilitation services to sick or injured persons who do not require hospitalization or institutionalization.
<b><u>URGENT CARE</u></b>	Urgent care facility: Location whose purpose is to diagnose and treat illness or injury for unscheduled, ambulatory patients seeking immediate medical attention. Emergency room—hospital: A portion of a hospital where emergency diagnosis and treatment of illness or injury is provided.
<b><u>OTHER</u></b>	All other places of service: Assisted living facility, birthing center, community mental health center, comprehensive inpatient rehabilitation facility, custodial care facility, end-stage renal disease treatment facility, federally qualified health center, group home, hospice, independent clinic, inpatient psychiatric facility, mass immunization center, military treatment facility, mobile unit, nursing facility, other place of service, psychiatric facility, psychiatric residential treatment center, rural health clinic, skilled nursing facility, public health clinic, unassigned, unknown, tribal 638 provider-based facility.

## 1.5 Grouping of Specialty

Specialties were grouped in consultation with hospital physicians, and informed by the specialty definitions provided by the American Medical Association and the Royal College of Physicians and Surgeons of Canada.

<b>SPECIALTY GROUP</b>	<b>DESCRIPTION</b>
<b><u>ANESTHESIOLOGY</u></b>	Anesthesiology, hyperbaric oxygen treatment, pain management.
<b><u>DIAGNOSTIC IMAGING</u></b>	Nuclear medicine, nuclear radiology, radiology, diagnostic radiology.
<b><u>EMERGENCY</u></b>	Emergency medicine, urgent care, intensivist, acute care.
<b><u>GENERAL PRACTICE</u></b>	Family practice, general practice.
<b><u>INTERNAL</u></b>	Allergy and immunology, cardiology, cardiology facility, cardiovascular disease, dermatology, dialysis center, endocrinology, metabolism, gastroenterology, geriatrics, gastrointestinal facility, hospital, infectious disease, internal medicine, nephrology, neurology, pulmonary disease, rheumatology, sleep medicine, sports medicine.
<b><u>LABORATORY</u></b>	Laboratory, reference laboratory.
<b><u>OBSTETRICS AND GYNECOLOGY</u></b>	Gynecological oncology, gynecology, obstetrics, reproductive endocrinology.
<b><u>PATHOLOGY</u></b>	Pathology, neuropathology
<b><u>PEDIATRIC</u></b>	Adolescent medicine, neonatology, pediatric allergy, cardiology, endocrinology, gastroenterology, hematology, oncology, infectious disease, nephrology, neurology, otolaryngology, pathology, pulmonology, radiology, rheumatology, surgery, and urology, perinatology.
<b><u>REHABILITATION</u></b>	Occupational medicine, orthotics and prosthetics, physical medicine and rehabilitation, physical therapy, rehabilitation therapy.
<b><u>SURGERY</u></b>	Abdominal, ambulatory, cardiovascular, colon and rectal, general, hand, head and neck, maxillofacial, neurological, neurosurgery, ophthalmology, orthopedic, outpatient, plastic, thoracic, trauma, urology, vascular, wound care.
<b><u>OTHER</u></b>	All other specialties: Acupuncture, ambulance, blood, chiropractic, clinical and social worker, clinical pharmacy, convalescent care, custodial care, dentist, durable medical equipment, genetics, hematology, home, hospice, infusion, marriage and family counseling, mental health, neuropsychology, not specified, nutrition, oncology, optomology, other, pharmacy, psychiatry, psychology, podiatry, public health and general medicine, radiation oncology, registered dietician, skilled nursing, speech therapy, transplant.

## 2 References

1. Charlson ME, Pompei P, Ales KL, and MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*. 2008; 40 (5): 373-383.
2. Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, and Kipnis P. Risk-Adjusting Hospital Inpatient Mortality Using Automated Inpatient, Outpatient, and Laboratory Databases. *Medical Care*. 2008; 46 (3): 232-239.
3. Gordy TR. CPT 2006 Current Procedural Terminology. 2006; American Medical Association.
4. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, and Ghali WA. Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Medical Care*. 2005; 43 (11): 1130-1139.