Appendix: Towards Assigning Diagnosis Codes using Medication History

Tomer Sagi[1][0000−0002−8916−0128], Emil Riis Hansen[1][0000−0003−4103−1244] Katja Hose[1][0000−0001−7025−8099], Gregory Y. H. Lip[2,5][0000−0002−7566−1626], Torben Bjerregaard Larsen[2,4][0000−0002−8485−8674], and Flemming Skjøth[2,3][0000−0003−2975−647X]

[1] Department of Computer Science, Aalborg University, Aalborg, Denmark
{tsagi,emilrh,khose}@cs.aau.dk
[2] Aalborg Thrombosis Research Unit, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
[3] Thrombosis and Drug Research Unit, Department of Research and Innovation, Aalborg University Hospital, Aalborg, Denmark
fls@rn.dk
[4] Thrombosis and Drug Research Unit, Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark
tobl@rn.dk
[5] Liverpool Centre for Cardiovascular Sciences, University of Liverpool and Liverpool Heart & Chest Hospital, Liverpool, United Kingdom
Gregory.Lip@liverpool.ac.uk

Appendix - Omitted codes and detailed results

Table 1 details the ommitted ocdes from the diagnosis table and the reasons for omission. We omit all codes with a low number of cases. We further omit 61 codes used to describe symptoms, as these are shared by multiple causes and will, most-probably, supplant a diagnosis code following medical investigation. Injuries and foreign bodies (30 codes) are omitted as well as their treatment is usually orthopedic or surgical, rather than medicinal. We omit the codes used in ICD-9 to classify birth-age and pre-term phase for infants (14 codes) as these are more descriptive than diagnostic. Finally, we omit the E and V series of codes that are used to provide additional details for statistical reasons and which do not cause differences in medicinal treatment. We remain with 567 codes and 54,423 cases (92.4%) that contain at least one of the remaining codes. Filtering out only admissions contained in both the diagnosis and prescription tables we remain with 50,211 admissions.

Tables 2 - 11 present the top and bottom F1-scoring codes and code groups at each aggregation level. The full results can be found online at **missing**.

Table 1. List of Omitted ICD-9 Codes and Code Groups

Code(s)	Description	Reason
5994 different codes	A large collection of various codes	Low base rate (less than 100 cases)
765.X	Descriptive of gestation week or preterm weight	Will be accompanied by the specific results of pre-term birth if such exist
8XX and 9XX	Injury	Medical result would be Surgical or Orthopedic and impossible to accurately specify from medication
93.31,93.41	Foreign body	Undiscernable medicinally
99.X	Complications of medical care	Undiscernable medicinally
61 different codes	Collection of different symptoms such as pain, nausea, and nuances of mental state/ faculties	Should be accompanied by the symptom's cause which is the main diagnosis

Table 2. MIMIC-III Diagnosis Prediction results for top 5 (F-1) ICD-9 level 0 codes

Codes	Description	Cases	Precision	Recall	F-1
760-779	Certain Conditions Originating In The Perinatal Period	277	99.3	99.3	99.3
390-459	Diseases Of The Circulatory System	3525	90.7	88.9	89.7
240-279	Endocrine, Nutritional And Metabolic Diseases, And Immunity Disorders	2856	80.7	69.1	74.5
580-629	Diseases Of The Genitourinary System	1906	77.9	56.0	65.2
460-519	Diseases Of The Respiratory System	1835	79.7	52.4	63.2

Table 3. MIMIC-III Diagnosis Prediction results for bottom 5 (F-1) ICD-9 level 0 codes

Codes	Description	Cases	Precision	Recall	F-1
680-709	Diseases Of The Skin And Subcutaneous Tissue	188	40.9	4.8	8.6
520-579	Diseases Of The Digestive System	888	69.0	6.5	11.9
290-319	Mental Disorders	799	82.8	9.0	16.3
740-759	Congenital Anomalies	131	68.4	9.9	17.3
001-139	Infectious And Parasitic Diseases	376	69.4	11.4	19.6

Table 4. MIMIC-III Diagnosis Prediction results for top 10 (F-1) ICD-9 level 1 codes

Codes	Description	Cases	Precision	Recall	F-1
764-779	Other Conditions Originating In The Perinatal Period	282	97.9	98.2	98.1
420-429	Other Forms Of Heart Disease	2,331	86.0	67.7	75.8
401-405	Hypertensive Disease	2,623	77.6	66.5	71.6
249-259	Diseases Of Other Endocrine Glands	1,318	83.0	61.0	70.3
580-589	Nephritis, Nephrotic Syndrome, and Nephrosis	1,402	74.1	55.5	63.5
042-042	Human Immunodeficiency Virus	58	76.5	44.8	56.5
270-279	Other Metabolic Disorders And Immunity Disorders	2,320	69.4	45.9	55.3
570-579	Other Diseases Of Digestive System	473	74.7	41.9	53.7
510-519	Other Diseases Of Respiratory System	1,112	69.2	41.9	52.2
410-414	Ischemic Heart Disease	856	72.5	39.1	50.8

Table 5. MIMIC-III Diagnosis Prediction results for bottom (F-1) ICD-9 level 1 codes

Codes	Description	Cases	Precision	Recall	F-1
720-724	Dorsopathies	99	0.0	0.0	0.0
179-189	Malignant Neoplasm Of Genitourinary Organs	38	0.0	0.0	0.0
260-269	Nutritional Deficiencies	167	14.3	0.6	1.1
393-398	Chronic Rheumatic Heart Disease	147	33.3	0.7	1.3
730-739	Osteopathies, Chondropathies, And Acquired Musculoskeletal Deformities	259	50.0	1.2	2.3
555-558	Noninfective Enteritis And Colitis	136	33.3	2.9	5.4
160-165	Malignant Neoplasm Of Respiratory And Intrathoracic Organs	103	60.0	2.9	5.6
710-719	Arthropathies And Related Disorders	197	46.7	3.6	6.6
680-686	Infections Of Skin And Subcutaneous Tissue	147	38.9	4.8	8.5
150-159	Malignant Neoplasm Of Digestive Organs And Peritoneum	36	22.2	5.6	8.9

Table 6. MIMIC-III Diagnosis Prediction results for top 10 (F-1) ICD-9 level 2 codes

Code	Description	Cases	Precision	Recall	F-1
774	Other perinatal jaundice	200	81.4	92.0	86.4
770	Other respiratory conditions of fetus and newborn	153	70.4	62.1	66.0
244	Acquired hypothyroidism	461	77.2	54.2	63.7
427	Cardiac dysrhythmias	1597	78.7	51.2	62.1
250	Diabetes mellitus	1236	79.8	50.2	61.6
414	Other forms of chronic ischemic heart disease	1274	83.5	46.9	60.0
042	Human immunodeficiency virus [HIV] disease	51	69.2	52.9	60.0
571	Chronic liver disease and cirrhosis	289	85.0	45.0	58.8
769	Respiratory distress syndrome in newborn	112	75.0	48.2	58.7
401	Essential hypertension	2000	64.6	51.0	57.0

Table 7. MIMIC-III Diagnosis Prediction results for bottom 10 (F-1) ICD-9 level 2 codes

Code	Description	Cases	Precision	Recall	F-1
135	Sarcoidosis	22	0.0	0.0	0.0
196	Secondary and unspecified malignant neoplasm of lymph nodes	70	0.0	0.0	0.0
202	Other malignant neoplasms of lymphoid and histiocytic tissue	22	0.0	0.0	0.0
251	Other disorders of pancreatic internal secretion	21	0.0	0.0	0.0
253	Disorders of the pituitary gland and its hypothalamic control	55	0.0	0.0	0.0
288	Diseases of white blood cells	129	0.0	0.0	0.0
289	Other diseases of blood and blood-forming organs	37	0.0	0.0	0.0
293	Transient mental disorders due to conditions classified elsewhere	142	0.0	0.0	0.0
300	Anxiety, dissociative and somatoform disorders	113	0.0	0.0	0.0
319	Unspecified intellectual disabilities	18	0.0	0.0	0.0

Table 8. MIMIC-III Diagnosis Prediction results for top 10 number (F-1) ICD-9 level 3 codes

Code	Description	Cases	Precision	Recall	F-1
774.2	Neonatal jaundice associated with preterm delivery	185	86.5	72.4	78.8
769	Respiratory distress syndrome in newborn	120	85.2	62.5	72.1
250.1	Diabetes with ketoacidosis	69	80.0	63.8	71.0
427.3	Atrial fibrillation and flutter	1225	76.4	59.7	67.0
585.6	End stage renal disease	194	82.3	55.2	66.0
414.0	Coronary atherosclerosis	1275	78.3	55.2	64.8
430	Subarachnoid hemorrhage	67	86.8	49.3	62.9
770.8	Other respiratory problems after birth	163	87.8	48.5	62.5
244.9	Unspecified acquired hypothyroidism	462	72.3	54.1	61.9
428.0	Congestive heart failure, unspecified	1243	68.1	52.3	59.2

Table 9. MIMIC-III Diagnosis Prediction results for bottom 10 (F-1) expected ICD-9 level 3 codes

Code	Description	Cases	Precision	Recall	F-1
038.8	Other specified septicemias	21	0.0	0.0	0.0
070.4	Other specified viral hepatitis with hepatic coma	20	0.0	0.0	0.0
112.5	Disseminated candidiasis	17	0.0	0.0	0.0
117.3	Aspergillosis	14	0.0	0.0	0.0
157.0	Malignant neoplasm of head of pancreas	8	0.0	0.0	0.0
162.8	Malignant neoplasm of other parts of bronchus or lung	31	0.0	0.0	0.0
185	Malignant neoplasm of prostate	17	0.0	0.0	0.0
196.1	Secondary and unspecified malignant neoplasm of intrathoracic lymph nodes	17	0.0	0.0	0.0
197.6	Secondary malignant neoplasm of retroperitoneum and peritoneum	31	0.0	0.0	0.0
211.3	Benign neoplasm of colon	40	0.0	0.0	0.0

Table 10. MIMIC-III Diagnosis Prediction results for top 10 number (F-1) ICD-9 level 4 codes

Code	Description	Cases	Precision	Recall	F-1
774.2	Neonatal jaundice associated with preterm delivery	179	79.3	85.5	82.3
244.9	Unspecified acquired hypothyroidism	467	70.0	65.3	67.6
427.31	Atrial fibrillation	1,197	72.4	62.1	66.8
769	Respiratory distress syndrome in newborn	122	81.3	53.3	64.4
414.01	Coronary atherosclerosis of native coronary artery	1,125	72.8	55.0	62.7
770.81	Primary apnea of newborn	102	65.2	58.8	61.9
585.6	End stage renal disease	193	68.7	52.3	59.4
430	Subarachnoid hemorrhage	50	68.4	52.0	59.1
428.0	Congestive heart failure, unspecified	1,198	63.5	51.9	57.1
571.2	Alcoholic cirrhosis of liver	121	72.2	47.1	57.0

Table 11. MIMIC-III Diagnosis Prediction results for bottom 10 (F-1) ICD-9 level 4 codes

Code	Description	Cases	Precision	Recall	F-1
778.4	Other disturbances of temperature regulation of newborn	10	0.0	0.0	0.0
771.81	Septicemia [sepsis] of newborn	12	0.0	0.0	0.0
770.82	Other apnea of newborn	13	0.0	0.0	0.0
770.2	Interstitial emphysema and related conditions	12	0.0	0.0	0.0
758.0	Down's syndrome	6	0.0	0.0	0.0
745.4	Ventricular septal defect	8	0.0	0.0	0.0
737.30	Scoliosis [and kyphoscoliosis], idiopathic	10	0.0	0.0	0.0
733.90	Disorder of bone and cartilage, unspecified	41	0.0	0.0	0.0
730.07	Acute osteomyelitis, ankle and foot	6	0.0	0.0	0.0
729.1	Myalgia and myositis, unspecified	37	0.0	0.0	0.0