Supplementary materials

**Supplementary Table 1.** A full list of baseline characteristics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Training cohort | Test cohort | *p*-value | Total |
| **Sociodemographic characteristics** |
| Age (years) | 54.3 (9.28)  | 94.27 (9.28)  | 0.43 | 54.29 (9.28) |
| Sex | Female | 42.2% (98048/232186)  |  42.2% (41987/99508) | 0.86 |  42.2% (140035/331694) |
| Male | 57.8% (134138/232186)  |  57.8% (57521/99508) |  |  57.8% (191659/331694) |
| Residence | Metropolitan | 16.7% (38808/232186)  |  16.5% (16371/995038) | 0.15 |  16.6% (55179/331694) |
| Urban or suburban | 28.2% (65516/232186)  |  28.2% (28066/99508) |  |  28.2% (93582/331694) |
| Rural | 55.1% (127862/232186)  |  55.3% (55071/99508) |  |  55.2% (182933/331694) |
| Income level | <30% | 28.6% (66305/232186)  |  28.4% (28309/99508) | 0.77 |  28.5% (94614/331694) |
| 30-80% | 34.9% (80925/232186)  |  35% (34788/99508) |  |  34.9% (115713/331694) |
| >80% | 36.6% (84956/232186)  |  36.6% (36411/99508) |  |  36.6% (121367/331694) |
| **Physical examination** |
| BMI (kg/m2) | 23.93 (2.88)  |  23.94 (2.89) | 0.41 |  23.93 (2.89) |
| Obesity  | Normal (BMI< 25 kg/m2) | 66.2% (153721/232070) | 66.1% (65756/99459) | 0.9 | 66.2% (219447/331529) |
| Overweight (25-30 kg/m2) | 31.3% (72675/323070) | 31.4% (31247/99459) |  | 31.3% (103922/331529) |
| Obese (> 30 kg/m2) | 2.4% (5674/232070) | 2.5% (2456/99459) |  | 2.5% (8130/331529) |
| Systolic BP (mmHg) | 126.57 (17.17)  |  126.63 (17.23) | 0.39 |  126.59 (17.19) |
| Diastolic BP (mmHg) | 79.18 (11.15)  |  79.18 (11.16) | 0.96 |  79.18 (11.15) |
| **Blood test** |
| AST (IU/L) | 26.56 (16)  | 26.52 (15.72)  | 0.42 | 26.55 (15.92)  |
| ALT (IU/L) | 25.52 (19.56)  | 25.50 (19.18)  | 0.77 | 25.52 (19.45) |
| GGT (IU/L) | 37.73 (52.2) | 37.64 (51.03) | 0.66 | 37.7 (51.85) |
| Total cholesterol (mg/dL) | 198.37 (36.86) | 198.2 (36.8) | 0.2 | 198.32 (36.84) |
| Fasting blood glucose (mg/dL) | 97.89 (29.18)  |  97.81 (28.24) | 0.42 |  97.87 (28.9) |
| Hemoglobin (g/dL) | 13.95 (1.49)  |  13.94 (1.49) | 0.73 |  13.94 (1.49) |
| **Urine stick test** |
| Urine pH level (pH) | 6.09 (0.64) |  6.1 (0.64) | 0.39 |  6.09 (0.64) |
| Urine Glucose | Negative or weak | 97.4% (225319/231397)  |  97.4% (96616/99173) | 0.8 |  97.4% (321935/330570) |
| Medium | 0.9% (2196/231397)  |  0.9% (941/99173) |  |  0.9% (3137/330570) |
| Strong | 1.7% (3882/231397) |  1.6% (1616/99173) |  |  1.7% (5498/330570) |
| Urine Occult blood | Negative or weak | 93.2% (215604/231385) | 93.1% (92285/99169) | 0.64 | 93.1% (307889/330554) |
| Medium |  3.8% (8776/231385) |  3.9% (3838/99169) |  |  3.8% (12614/330554) |
| Strong |  3% (7005/231385) |  3.1% (3046/99169) |  |  3% (10051/330554) |
| Urine protein | Negative or weak |  98.2% (227156/231402) |  98.1% (97326/99177) | 0.24 |  98.2% (324482/330579) |
| Medium |  1.3% (2903/231402) |  1.2% (1219/99177) |  |  1.2% (4122/330579) |
| Strong |  0.6% (1343/231402) |  0.6% (632/99177) |  |  0.6% (1975/3305/79) |
| **Habitus** |
| Smoking (pack-year) |  5.98 (11.56)  |  5.92 (11.46) | 0.18 |  5.96 (11.53) |
| Alcohol consumption | Rarely |  56.3% (128639/2283638) |  56.2% (590593/9/8/35) | 0.48 |  56.3% (183692/326241) |
| 2-3 per month |  15.6% (35665/228368) |  15.5% (15180/97873) |  |  15.6% (50845/326241) |
| 1-2 per week |  17.4% (39629/228368) |  17.5% (17135/97873) |  |  17.4% (56764/326241) |
| 3-4 per week |  6.8% (15453/228368) |  6.9% (6744/97873) |  |  6.8% (22197/326241) |
| Almost everyday |  3.9% (8982/228368) |  3.8% (3761/97873) |  |  3.9% (12743/326241) |
| Exercise | Rarely |  50.1% (113877/227/095) |  50.1% (48791/97411) | 0.59 |  50.1% (162668/324506) |
| 1-2 per week |  26.9% (61202/227095) |  26.9% (26241/9/7411) |  |  26.9% (87443/324506) |
| 3-4 per week |  12.1% (27560/22/095) |  12.1% (11781/97411) |  |  12.1% (39341/324506) |
| 5-6 per week |  3.2% (/2/0/227095) |  3.2% (3114/97411) |  |  3.2% (10384/324506) |
| Almost everyday |  7.6% (17186/227095) |  7.7% (7484/97411) |  |  7.6% (24670/324506) |
| **Family history** |
| Liver disease |  2.8% (5989/213528) |  2.81% (2574/91558) | 0.9 |  2.81% (8563/305086) |
| Hypertension |  9.16% (19648/214527/) |  9.16% (8424/92012) | 0.77 |  9.16% (28072/306539) |
| Stroke |  5.49% (11742/214048) |  5.44% (4996/91769) | 0.85 |  5.47% (16738/305817) |
| Heart disease |  2.39% (5102/213537) |  2.38% (21/76/91569) | 0.85 |  2.39% (7278/305106) |
| Diabetes mellitus |  6.4% (13718/214195) |  6.54% (6011/91853) | 0.3 |  6.45% (19729/306048) |
| Cancer | 13.1% (28189/215156) | 13.22% (12200/9228/) | 0.49 | 13.14% (40389/307443) |
| **Underlying medical condition** |
| Diabetes mellitus |  6.08% (14114/232186) |  6.05% (6025/99508) | 0.8 |  6.07% (20139/331694) |
| Dyslipidemia |  5.95% (13811/232186) |  5.94% (5914/99508) | 0.96 |  5.95% (19725/331694) |
| Hypertension |  19.08% (44298/232186) |  19.31% (19213/99508) | 0.13 |  19.15% (63511/331694) |
| Chronic hepatitis virus infection | 2.55% (5915/232186) | 2.63% (2615/99508) | 0.16 | 2.57% (8530/331694) |
| Human Immunodeficiency virus | 6.69% (15527/232186) | 6.82% (6787/99508) | 0.16 | 6.73% (22314/331694) |
| Schizophrenic or delusional disorders, or mental disorders due to psychoactive substance use | 15.55% (36110/232186) |  15.68% (15604/99508) | 0.35 |  15.59% (51714/331694) |
| Chronic liver disease |  5.77% (13397/232186) |  5.17% (5149/99508) | 0.15 |  5.59% (18546/331694) |
| Stroke |  0.02% (56/232186) |  0.02% (16/99508) | 0.19 |  0.02% (72/331694) |
| Ischemic heart disease |  2.68% (6232/232186) |  2.64% (2627/99508) | 0.48 |  2.6/% (8859/331694) |
| Atrial fibrillation |  0.41% (952/232186) |  0.4% (399/99508) | 0.73 |  0.41% (1351/331694) |
| Gastrointestinal bacterial infection |  11.55% (26818/232186) |  11.69% (11634/99508) | 0.25 |  11.59% (38452/331694) |
| Other bacterial infection |  5.10% (11920/232186) |  5.01% (4983/99508) | 0.74 |  5.10% (16903/331694) |
| Other viral infection |  6.22% (14434/232186) |  6.18% (6145/99508) | 0.66 |  6.2% (205/9/331694) |
| Fungal infection |  21.14% (49092/232186) |  21.38% (21294/99508) | 0.13 |  21.21% (70368/331694) |
| Gastrointestinal benign tumors |  1.51% (3500/232186) |  1.49% (1479/99508) | 0.66 |  1.5% (4979/331694) |
| Thyroid benign tumors |  0.41% (942/232186) |  0.44% (437/99508) | 0.18 |  0.42% (1379/331694) |
| Other benign tumors |  3.09% (7176/232186) |  3.16% (3147/99508) | 0.28 |  3.11% (10323/331694) |
| Blood disorders |  3.83% (8883/232186) |  3.81% (3791/99508) | 0.83 |  3.82% (126/4/331694) |
| Thyroid diseases |  4.2% (7176/232186) |  4.19% (4168/99508) | 0.85 |  4.2% (13928/331694) |
| Other endocrine diseases |  0.42% (964/232186) |  0.39% (393/995038) | 0.42 |  0.41% (135//331694) |
| Other metabolic diseases | 1.01% (2334/232186) | 1.01% (1007/99508) | 0.87 | 1.01% (3341/331694) |
| Mood disorders |  4.47% (10370/232186) |  4.51% (4484/99508) | 0.62 |  4.48% (14854/331694) |
| Mental disorder due to organic causes | 0.23% (531/232186) | 0.22% (221/99508) | 0.74 | 0.23% (752/331694) |
| Headache |  10.93% (25379/232186) |  10.8% (10751/99508) | 0.29 |  10.89% (36130/331694) |
| Sleep disorders |  2.36% (5479/232186) |  2.46% (2447/9950) | 0.09 |  2.39% (7926/331694) |
| Neuropathy |  3.96% (19878/232186) |  8.63% (8586/99508) | 0.53 |  8.58% (28464/331694) |
| Eye diseases |  45.05% (104610/232186) |  45.03% (44805/99508) | 0.89 |  45.05% (149415/331694) |
| Ear diseases |  20.89% (48513/232186) |  20.74% (20641/99508) | 0.33 |  20.85% (69154/331694) |
| Peripheral vascular diseases |  3.06% (7114/232186) |  3.04% (3021/99508) | 0.68 |  3.06% (10135/331694) |
| Varicose |  0.54% (1252/232186) |  0.54% (540/99508) | 0.92 |  0.54% (1792/331694) |
| Hemorrhoid |  6.13% (14237/232186) |  6.06% (6030/99508) | 0.43 |  6.11% (2026//331694) |
| Congestive heart failure |  1.28% (2973/232186) |  1.33% (1326/99508) | 0.23 |  1.3% (4299/331694) |
| Chronic obstructive pulmonary disease |  11.1% (25772/232186) |  11.03% (10974/99503) | 0.55 |  11.08% (36746/331694) |
| Asthma |  10.77% (25009/232186) |  10.81% (10758/99508) | 0.74 |  10.78% (35767/331694) |
| Bronchiectasis |  0.92% (2134/232186) |  0.88% (871/99508) | 0.23 |  0.91% (3005/331694) |
| Lung disease due to external causes |  0.06% (132/232186) |  0.07% (66/99508) | 0.34 |  0.06% (198/331694) |
| Allergic rhinitis |  30.04% (69750/232186) |  29.58% (29436/99508) | 0.01 |  29.9% (99186/331694) |
| Chronic rhinitis |  10.71% (24869/232186) |  10.57% (10520/99508) | 0.24 |  10.67% (35389/331694) |
| Sinusitis |  7.19% (16690/2321386) |  7.04% (7003/99508) | 0.12 |  7.14% (23693/331694) |
| Nasal polyp | 0.79% (1843/232186) | 0.78% (779/99508) | 0.76 | 0.79% (2622/331694) |
| Throat diseases |  8.96% (20793/232186) | 8.86% (8818/99508) | 0.39 |  8.93% (29611/331694) |
| Esophagitis |  14.2% (329/0/232186) | 13.99% (13921/99508) | 0.11 |  14.14% (46891/331694) |
| Gastritis |  96.53% (131248/232186) |  96.52% (56237/99508) | 0.95 |  56.52% (18/485/331694) |
| Inflammatory bowel disease |  0.78% (1822/232186) |  0.8% (799/99508) | 0.6 |  0.79% (2621/331694) |
| Irritable bowel syndrome |  17.13% (39784/232186) |  16.98% (16896/99508) | 0.28 |  17.09% (56680/331694) |
| Anal diseases |  1.49% (3449/232186) |  1.41% (1404/99508) | 0.1 |  1.46% (4853/331694) |
| Alcoholic liver disease |  3.2% (7427/232186) |  3.21% (3192/99508) | 0.9 |  3.2% (10619/3316094) |
| Toxic liver disease |  0.88% (2042/232186) |  0.87% (866/99508) | 0.81 |  0.88% (2908/331694) |
| Other liver diseases |  9.23% (21432/232186) |  9.11% (9065/99508) | 0.27 |  9.19% (30497/331694) |
| Cholecystitis |  0.26% (598/232186) |  0.28% (274/99508) | 0.38 |  0.26% (8/2/331694) |
| Other cholecystic diseases |  0.11% (259/232186) |  0.11% (107/99508) | 0.79 |  0.11% (366/331694) |
| Biliary diseases |  0.13% (293/232186) |  0.12% (124/99508) | 0.95 |  0.13% (417/331694) |
| Acute pancreatitis |  0.26% (612/232186) |  0.27% (265/99508) | 0.92 |  0.26% (877/331694) |
| Other pancreatic diseases |  0.21% (480/232186) |  0.2% (199/99508) | 0.72 |  0.2% (679/331694) |
| Skin diseases |  48.73% (113151/232186) |  43.96% (48326/99508) | 0.38 |  43.68% (161477/331694) |
| Musculoskeletal diseases |  65.19% (1513 /0/232186) |  64.87% (64546/99508) | 0.07 |  65.09% (215916/331694) |
| Urolithiasis |  2.29% (5220/232186) |  2.22% (2209/99508) | 0.62 |  2.24% (7429/331694) |
| Other urologic diseases |  20.71% (48077/232186) |  20.87% (20765/99508) | 0.3 |  20.75% (68842/331694) |

For continuous variables, numbers in parentheses are standard deviation. BMI=body mass index, BP=blood pressure, AST=aspartate aminotransferase, ALT=alanine aminotransferase, GGT=gamma-glutamyl transferase.

**Supplementary Table 2.** Operational definitions of underlying medical conditions

|  |  |
| --- | --- |
| Hepatocellular carcinoma | C22.0 AND admission |
| Other cancers | (C codes except for C22.0) AND admission |
| Diabetes mellitus | E11-14 AND antidiabetic medication |
| Dyslipidemia | E78 AND lipid-lowering medication |
| Hypertension | (I10-13, I15) AND (admission OR outpatient ≥2) |
| Chronic hepatitis virus infection | B15-19 |
| Human Immunodeficiency virus | B\_ |
| Schizophrenic or delusional disorders or mental and behavioral disorders due to psychoactive substance use | F\_ |
| Chronic liver disease | K72-74 |
| Stroke | I60-64 AND (admission OR emergency room) AND imaging study |
| Ischemic heart disease | I20-25 AND (admission OR death within 1 month due to cardiovascular diseases OR outpatient visits ≥ 4) |
| Atrial fibrillation | I48 AND (admission OR outpatient ≥2) |
| Gastrointestinal bacterial infection | A01-09 |
| Other bacterial infection | A20-99 |
| Other viral infection | B20-34 |
| Fungal infection | B35-49 |
| Gastrointestinal benign tumors | D12-13 |
| Thyroid benign tumors | D34 |
| Other benign tumors | D\_ |
| Blood disorders | D5-8 |
| Thyroid diseases | E0 |
| Other endocrine diseases | E2-3 |
| Other metabolic diseases | E79, E8, E9 |
| Mood disorders | F3 |
| Mental disorder due to organic causes | F0 |
| Headache | G43-44 |
| Sleep disorders | G47 |
| Neuropathy | G5-6 |
| Eye diseases | H0-5 |
| Ear diseases | H6-9 |
| Peripheral vascular diseases | I73 |
| Varicose | I83 |
| Hemorrhoid | I84 |
| Congestive heart failure | I50 |
| Chronic obstructive pulmonary disease | J41-44 |
| Asthma | J45-46 |
| Bronchiectasis | J47 |
| Lung disease due to external causes | J60-67 |
| Allergic rhinitis | J30 |
| Chronic rhinitis | J31 |
| Sinusitis | J32 |
| Nasal polyp | J33 |
| Throat diseases | J35-38 |
| Esophagitis | K20-21, K221, K227 |
| Gastritis | K25-29 |
| Inflammatory bowel disease | K50-51 |
| Irritable bowel syndrome | K58 |
| Anal diseases | K60-61 |
| Alcoholic liver disease | K70 |
| Toxic liver disease | K71 |
| Other liver diseases | K76-77 |
| Cholecystitis | K81 |
| Other cholecystic diseases | K82 |
| Biliary diseases | K83 |
| Acute pancreatitis | K85 |
| Other pancreatic diseases | K86 |
| Skin diseases | All L codes |
| Musculoskeletal diseases | All M codes |
| Chronic renal disease | N18-19 AND (admission OR outpatient ≥2) |
| Urolithiasis | N20-22 |
| Other urologic diseases | N\_ |

Codes used here are according to the Korean Standard Classification of Diseases version 6 ((KCD-6) which is based on the International Classification of Diseases 10th revision (ICD-10).

**Supplementary Table 3.** Frequency of each variable being selected as a significant predictor of hepatocellular carcinoma development in Cox proportional hazard regression using 1,000 bootstrapped training datasets. The variables in bold showed >95% frequencies.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Frequency** | **Variable** | **Frequency** |
| **Age** | 100% (1,000/1,000) | **Sex** | 100% (1,000/1,000) |
| **Chronic liver disease** | 100% (1,000/1,000) | **gamma-glutamyl transpeptidase** | 100% (1,000/1,000) |
| **Family history of liver disease** | 100% (1,000/1,000) | **Alanine aminotransferase** | 100% (1,000/1,000) |
| **Chronic hepatitis virus infection** | 100% (1,000/1,000) | **Total cholesterol** | 100% (1,000/1,000) |
| **Human Immunodeficiency virus** | 100% (1,000/1,000) | **Dyslipidemia** | 99.4% (994/1,000) |
| **Diabetes mellitus** | 98.7% (987/1,000) | **Schizophrenic or delusional disorders, or mental disorders due to psychoactive substance use** | 96.3% (963/1,000) |
| Drinking habit | 76.6% (766/1,000) | Esophagitis | 76.9% (769/1,000) |
| Fungal infection | 75.5% (755/1,000) | Chronic rhinitis | 74.2% (742/1,000) |
| Other liver diseases | 73.3% (733/1,000) | Hypertension | 66.0% (660/1,000) |
| Urine glucose | 63.2% (632/1,000) | Eye diseases | 61.3% (613/1,000) |
| Urine protein | 59.6% (596/1,000) | Obesity | 58.9% (589/1,000) |
| Hemorrhoid | 52.4% (524/1,000) | Ear diseases | 52.8% (528/1,000) |
| Income level | 50.1% (501/1,000) | Other benign tumors | 49.8% (498/1,000) |
| Alcoholic liver disease | 46.9% (469/1,000) | Thyroid diseases | 41.2% (412/1,000) |
| Headache | 40.4% (404/1,000) | Allergic rhinitis | 38.2% (382/1,000) |
| Cholecystitis | 37.3% (373/1,000) | Urolithiasis | 36.8% (368/1,000) |
| Irritable bowel syndrome | 29.6% (296/1,000) | Exercise habit | 28.4% (248/1,000) |
| Other bacterial infection | 28.1% (281/1,000) | Lung disease due to external agents | 21.8% (218/1,000) |
| Gastrointestinal bacterial infection | 21.5% (215/1,000) | Other urologic diseases | 20.1% (201/1,000) |
| Other gallbladder diseases | 18.1% (181/1,000) | Gastritis | 17.0% (170/1,000) |
| Skin diseases | 15.8% (158/1,000) | Hemoglobin | 13.1% (131/1,000) |
| Family history of diabetes mellitus | 12.1% (121/1,000) | Fasting blood glucose level | 11.9% (119/1,000) |
| Family history of heart disease | 11.2% (112/1,000) | Mood disorders | 10.8% (108/1,000) |
| Residence | 10.7% (107/1,000) | Benign tumors in lower gastrointestinal tract | 10.0% (100/1,000) |
| Diseases in anus | 9.8% (98/1,000) | Acute pancreatitis | 9.6% (96/1,000) |
| Diseases in throat | 9.3% (93/1,000) | Other viral infection | 9.3% (93/1,000) |
| Varicose | 8.2% (82/1,000) | Other bacterial infection | 7.9% (79/1,000) |
| Other metabolic diseases | 7.9% (79/1,000) | Chronic obstructive pulmonary disease | 7.5% (75/1,000) |
| Other biliary diseases | 7.2% (72/1,000) | Hypertension | 7.2% (72/1,000) |
| Urine pH | 7.2% (72/1,000) | Musculoskeletal diseases | 6.2% (62/1,000) |
| Urine occult blood | 6.0% (60/1,000) | Smoking | 5.9% (59/1,000) |
| Nasal polyps | 5.6% (56/1,000) | Family history of stroke | 5.4% (54/1,000) |
| Toxic liver disease | 5.3% (53/1,000) | Other pancreatic diseases | 5.1% (50/1,000) |
| Neuropathy | 4.8% (48/1,000) | Blood disorders | 4.6% (46/1,000) |
| Disability | 4.0% (40/1,000) | Bronchiectasis | 3.5% (35/1,000) |
| Sinusitis | 3.1% (31/1,000) | Family history of cancer | 2.7% (27/1,000) |
| Ischemic heart disease | 2.7% (27/1,000) | Sleep disorders | 2.7% (27/1,000) |
| Asthma | 2.6% (26/1,000) | Atrial fibrillation | 2.1% (21/1,000) |