

Introduction

心衰和心梗【所有入院首诊为心衰心梗的病人，不限于ICU，若多次因为首诊心衰或者心梗入院，则提取第一次入院信息】：

- 1. 人口学基本信息和共病情况 data_info4
- 2. 实验室指标 data_chart & data_lab
- 3. 用药信息 data_pharmacy
- 4. 在icu停留的记录 data_icustay

Tasks

- 1. 把实验室指标、用药信息 长数据转成宽数据
- 2. 把上述表格合并成一个宽数据的表格

```
Out[2]: 'E:\\Experiments\\Lei.Zhang\\01-Working\\Datathon2022\\metacsv'

D:\Dev\anaconda3\envs\pytorch\lib\site-packages\IPython\core\interactiveshell.py:3457: DtypeWarning: Columns (5) have mixed
types.Specify dtype option on import or set low_memory=False.
  exec(code_obj, self.user_global_ns, self.user_ns)
```

随机展示

Chart表, 实验室表, 以及用药表的分析数据有:18504552, 3197442, 667606
Chart表、实验表、用药表中的 住院号数量: 4297, 12257, 12352
Chart表、实验表、用药表中的 病人数量: 4297, 12257, 12352

基本信息

Out[5]:

	subject_id	hadm_id	admittime	disctime	admission_type	admission_location	discharge_location	insurance	language	marital_status	...	peptic
12423	19996783	21880161	9/5/2188 15:55:00	19/5/2188 18:09:00	OBSERVATION ADMIT	WALK-IN/SELF REFERRAL	HOSPICE	Other	?	MARRIED	...	
1889	11508828	28435746	5/3/2145 18:15:00	16/3/2145 16:57:00	URGENT	TRANSFER FROM HOSPITAL	CHRONIC/LONG TERM ACUTE CARE	Medicare	ENGLISH	MARRIED	...	
7797	16183439	28323401	7/6/2131 15:54:00	12/6/2131 16:10:00	EW EMER.	PHYSICIAN REFERRAL	HOME	Other	ENGLISH	MARRIED	...	
9426	17529599	20817775	16/9/2171 16:38:00	18/9/2171 15:00:00	URGENT	TRANSFER FROM HOSPITAL	HOME	Other	ENGLISH	MARRIED	...	
149	10131753	22565233	11/6/2125 15:52:00	13/6/2125 13:43:00	DIRECT EMER.	PHYSICIAN REFERRAL	HOME	Medicare	ENGLISH	MARRIED	...	
539	10422699	26593792	14/2/2143 21:11:00	18/2/2143 12:44:00	EW EMER.	EMERGENCY ROOM	HOME	Medicare	ENGLISH	MARRIED	...	
3402	12698808	24858208	23/9/2182 23:05:00	25/9/2182 14:29:00	URGENT	TRANSFER FROM HOSPITAL	HOME	Other	ENGLISH	SINGLE	...	
1714	11352800	28431901	2/3/2134 17:50:00	7/3/2134 15:40:00	EW EMER.	EMERGENCY ROOM	HOME HEALTH CARE	Medicare	ENGLISH	MARRIED	...	
12075	19699260	25530837	5/3/2124 13:33:00	6/3/2124 16:58:00	EW EMER.	EMERGENCY ROOM	SKILLED NURSING FACILITY	Other	ENGLISH	WIDOWED	...	
3208	12556596	29031376	23/2/2120 17:28:00	1/3/2120 15:56:00	OBSERVATION ADMIT	EMERGENCY ROOM	HOME HEALTH CARE	Medicare	ENGLISH	MARRIED	...	

10 rows × 31 columns

患有‘心衰’/‘心梗’的病人:12428

基本信息表的每一行代表一个主体: True

基本信息表的所有列信息
'subject_id', 'hadm_id', 'admittime', 'disctime', 'admission_type', 'admission_location', 'discharge_location', 'insurance', 'language', 'marital_status', 'ethnicity', 'status', 'gender', 'age', 'myocardial_infarct', 'congestive_heart_failure', 'peripheral_vascular_disease', 'cerebrovascular_disease', 'dementia', 'chronic_pulmonary_disease', 'rheumatic_disease', 'peptic_ulcer_disease', 'mild_liver_disease', 'diabetes_without_cc', 'diabetes_with_cc', 'paraplegia', 'renal_disease', 'malignant_cancer', 'severe_liver_disease', 'metastatic_solid_tumor', 'aids'

基本信息表中关键字

info4 更多 ([https://mimic.mit.edu/docs/iv/modules/hosp/admissions/%5D\(https://mimic.mit.edu/docs/iv/modules/icu/chartevents/\)](https://mimic.mit.edu/docs/iv/modules/hosp/admissions/%5D(https://mimic.mit.edu/docs/iv/modules/icu/chartevents/))): 有关住院的详细信息。

入院表提供了关于病人入院的信息。由于病人的每一次独特的医院访问都被分配了一个独特的hadm_id，所以入院表可以被认为是hadm_id的定义表。可用的信息包括入院和出院的时间信息、人口统计信息、入院的来源等等。

- 数据来自医院的入院、出院和转院数据库（通常称为“ADT”数据）。
- 有时会为在医院死亡的患者创建器官捐献者账户。这些是不同的住院时间，住院时间很短，有时甚至是负数。此外，他们的死亡时间通常与较早入院患者的死亡时间相同。

- subject_id, hadm_id : 该表的每一行都包含一个唯一的 hadm_id，代表单个患者入院。 hadm_id的取值范围是2000000-2999999， 这张表subject_id可能重复，说明一个病人多次入院。可以使用 subject_id 将 ADMISSIONS 表链接到 PATIENTS 表。
- admittime, disctime, (入院时间、离院时间、死亡时间): admittime提供病人入院的日期和时间，disctime则提供病人出院的日期和时间。如果适用，deathtime提供病人在医院内死亡的时间。请注意，只有当病人在医院内死亡时才会出现deathtime，而且几乎总是与病人的disctime相同。然而，由于打字错误，可能会有一些差异。

💡 deathtime 无

- admission_type : 用于对入院的紧迫性进行分类。有9种可能性。门诊观察"、"直接急诊"、"直接观察"、"择期"、"紧急观察"、"观察入院"、"当天手术入院"、"紧急"。
- admission_location : 提供病人到达医院之前的位置信息。请注意，由于急诊室在技术上是一个诊所，通过急诊室入院的病人通常将其作为入院地点。
- 同样，discharge_location 是患者出院后的处置情况。
- insurance, language, marital_status, ethnicity (保险、语言、婚姻状况、种族): 保险、语言、婚姻状况和种族列提供有关给定住院的患者人口统计信息。请注意，由于每次入院都会记录此数据，因此它们可能会因住院而异。

💡 edregtime, edouttime (暂无) 患者在急诊室登记和出院的日期和时间。



向外链接-Links to:

- patients on subject_id
- admissions on hadm_id
- icustays on stay_id
- d_items on itemid

基本信息表初始统计: ‘心衰’/‘心梗’患者

【字典】疾病的编码 (表Infor4)

- subject_id
- status :
- gender :
- age :

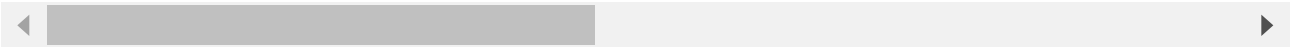
- myocardial_infarct : 心肌梗死 
- congestive_heart_failure : 充血性心力衰竭' 
- peripheral_vascular_disease : 周边血管疾病
- cerebrovascular_disease : 脑血管疾病
- dementia : 痴呆症
- chronic_pulmonary_disease : 慢性肺部疾病
- rheumatic_disease : 风湿病
- peptic_ulcer_disease : 消化性溃疡病
- mild_liver_disease : 轻度肝脏疾病
- diabetes_without_cc : 无cc的糖尿病
- diabetes_with_cc : 有cc的糖尿病
- paraplegia : 截瘫
- renal_disease : 肾脏疾病
- malignant_cancer : 恶性肿瘤
- severe_liver_disease : 严重的肝脏疾病
- metastatic_solid_tumor : 转移性固体肿瘤
- aids : 艾滋病

统计基本信息表中的疾病分布

Out[8]:

	myocardial_infarct	congestive_heart_failure	peripheral_vascular_disease	cerebrovascular_disease	dementia	chronic_pulmonary_disease	rheumatic_
0	0	1	0	0	0		1
1	1	0	0	0	0		0
2	1	1	0	0	0		0
3	1	1	0	0	0		1
4	1	1	0	0	0		0
...
12423	1	0	0	0	0		0
12424	1	1	0	0	0		1
12425	1	1	0	0	0		0
12426	0	1	0	0	0		1
12427	1	0	0	0	0		0

12428 rows × 17 columns

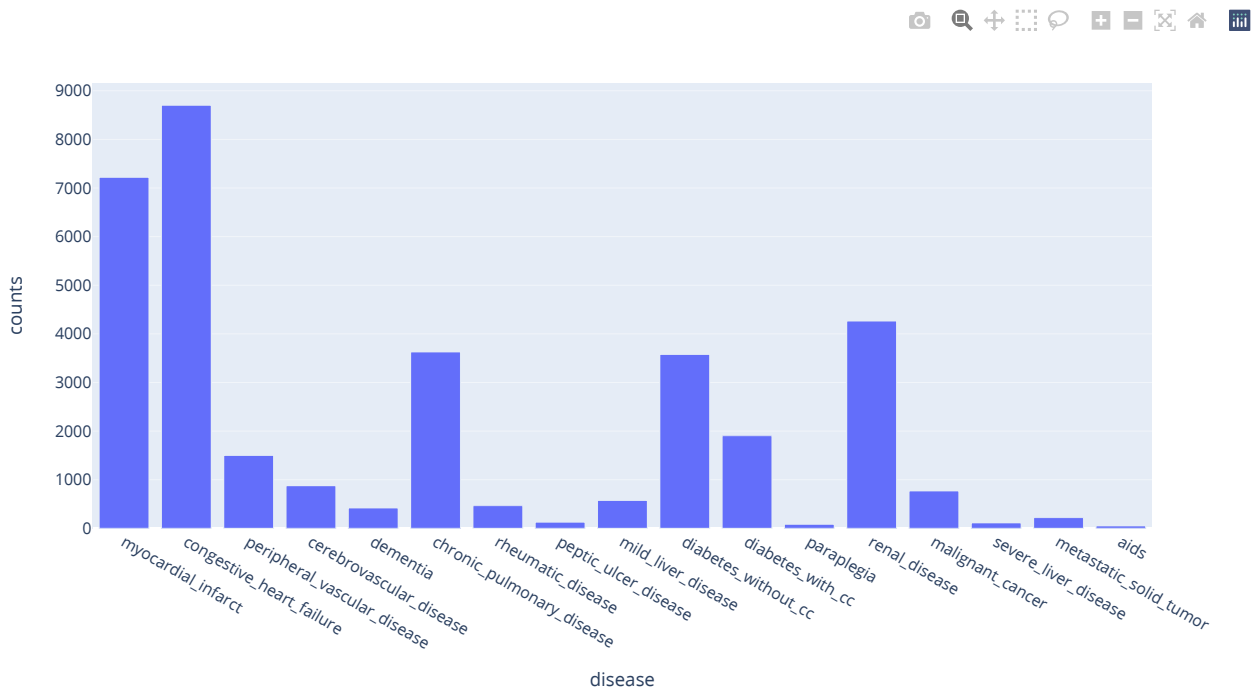


降序排列疾病

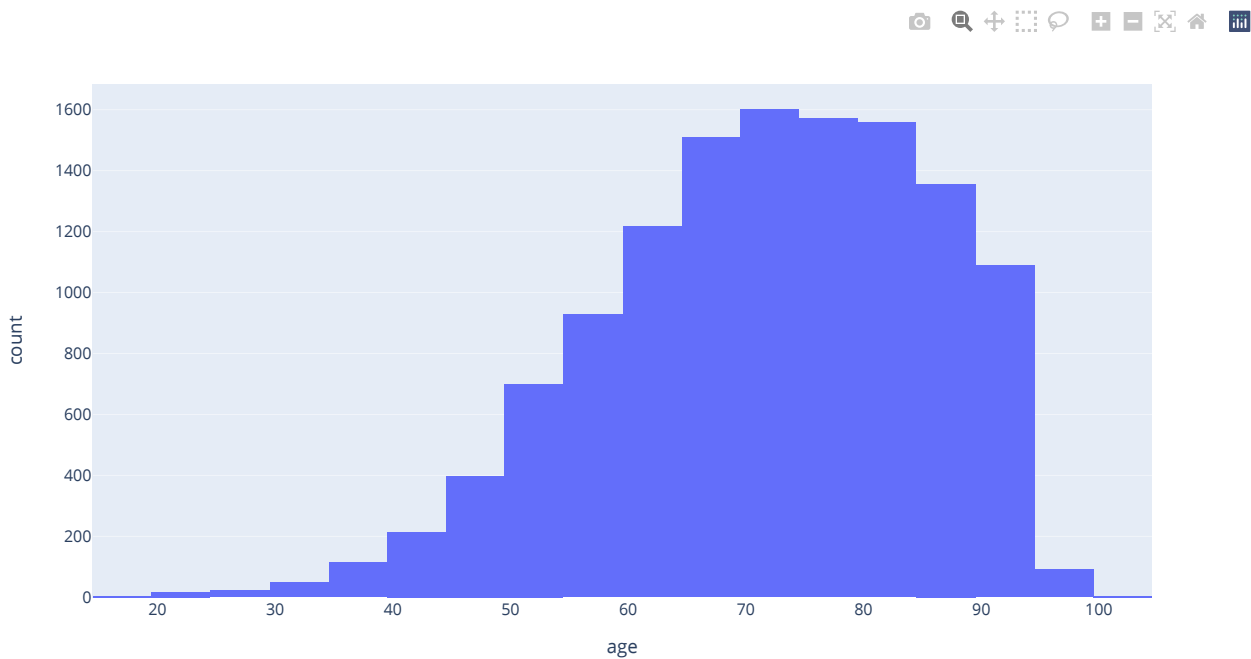
Out[9]:

	disease	counts
0	congestive_heart_failure	8702
1	myocardial_infarct	7220
2	renal_disease	4265
3	chronic_pulmonary_disease	3629
4	diabetes_without_cc	3577
5	diabetes_with_cc	1906
6	peripheral_vascular_disease	1500
7	cerebrovascular_disease	877
8	malignant_cancer	770
9	mild_liver_disease	574
10	rheumatic_disease	467
11	dementia	420
12	metastatic_solid_tumor	221
13	peptic_ulcer_disease	127
14	severe_liver_disease	113
15	paraplegia	83
16	aids	51

分布图



基本信息表中的年龄分布



同时心衰&心梗

Chart表

Out[12]:

	subject_id	hadm_id	charttime		storetime		itemid	value	valuenum
0	10464977	23034003	7/11/2140	20:58:00	7/11/2140	20:58:00	223956	Ventricular Demand	NaN
1	10464977	23034003	7/11/2140	20:58:00	7/11/2140	20:58:00	223961	Attached-Pacer	NaN
2	10464977	23034003	7/11/2140	20:58:00	7/11/2140	20:58:00	224390	No	NaN
3	10464977	23034003	7/11/2140	20:58:00	7/11/2140	20:58:00	224834	Yes	NaN
4	10464977	23034003	7/11/2140	20:58:00	7/11/2140	20:58:00	224835	Yes	NaN
...
18504547	10464977	23034003	3/11/2140	00:00:00	3/11/2140	03:12:00	227947	Confusion	NaN
18504548	10464977	23034003	3/11/2140	00:00:00	3/11/2140	03:12:00	227948	Soft limb	NaN
18504549	10464977	23034003	3/11/2140	00:00:00	3/11/2140	03:12:00	227949	Done	NaN
18504550	10464977	23034003	3/11/2140	00:00:00	3/11/2140	03:12:00	227950	Both arms	NaN
18504551	10464977	23034003	3/11/2140	00:00:00	3/11/2140	03:12:00	227951	Circulation adequate	NaN

18504552 rows × 7 columns

chart表条目:18504552

chart表统计subject数量: 4297, 平均每个subject进行的ICU项目: 4306

Out[14]: 18504552

1.1 ICU表中字段解释

Chartevent 更多 (https://mimic.mit.edu/docs/iv/modules/icu/chartevents/): 在 ICU 逗留期间发生的图表项目。包含 ICU 中记录的大部分信息。

chartevents 包含所有可用于患者的图表数据。在他们入住 ICU 期间，患者信息的主要存储库是他们的电子图表。电子图表显示患者的常规生命体征和与他们的护理相关的任何其他信息：**呼吸机设置、实验室值、代码状态、精神状态**等。因此，有关患者住院的大部分信息都包含在图表事件中。此外，即使其他地方 (labevents) 捕获了实验室值，它们也经常 在 chartevents 中重复。发生这种情况是因为希望在患者的电子图表上显示实验室值，因此将值从存储实验室值的数据库复制到存储图表事件的数据库。

- 指定患者的标识符： subject_id 对于患者是唯一的， hadm_id 对于患者住院是唯一的， stay_id 对于患者住院是唯一的。有关这些标识符的更多信息，请参见此处。
- charttime 记录进行观察的时间，通常是最接近实际测量数据的时间。 storetime 记录临床工作人员手动输入或手动验证观察结果的时间。
- 数据库中单一测量类型的标识符。与一个 ITEMID （如212）相关的每一行都对应于同一测量（如心率）的一个实例。
- value 包含为 ITEMID 标识的概念测量的值。如果此值为数字，则 valuenum 包含数字格式的相同数据。如果此数据不是数字，则 valuenum 为空。在某些情况下（例如 Glasgow Coma Scale、Richmond Sedation Agitation Scale 和 Code Status 等分数）， valuenum 包含分数， value 包含分数和描述分数含义的文本。

Links to:

- patients on subject_id
- admissions on hadm_id
- icustays on stay_id
- d_items on itemid

Lab表

Out[15]:

	subject_id	hadm_id	charttime	storetime	itemid	value	valuenum
0	10013653	26666796	26/10/2182 07:40:00	26/10/2182 09:02:00	50960	2.0	2.00
1	10013653	26666796	26/10/2182 00:57:00	26/10/2182 01:02:00	50802	-2	-2.00
2	10013653	26666796	26/10/2182 00:57:00	26/10/2182 01:02:00	50804	23	23.00
3	10013653	26666796	26/10/2182 00:57:00	26/10/2182 01:02:00	50818	37	37.00
4	10013653	26666796	26/10/2182 00:57:00	26/10/2182 01:02:00	50820	7.38	7.38
...
3197437	16221250	23282331	26/9/2140 22:22:00	26/9/2140 22:50:00	51508	Red	NaN
3197438	12741325	20860534	26/1/2121 12:35:00	26/1/2121 14:16:00	51508	PINK	NaN
3197439	19119896	28318588	2/4/2195 16:00:00	2/4/2195 16:54:00	51508	Red	NaN
3197440	19506871	21026986	11/4/2118 22:50:00	11/4/2118 23:30:00	51508	DKAMB	NaN
3197441	18426598	22940206	24/4/2173 01:30:00	24/4/2173 02:38:00	51508		NaN

3197442 rows × 7 columns

Lab表条目:3197442

Lab表统计subject数量: 12257, 平均每个subject进行的Lab项目: 260

实验室表中字段解释

labevents (https://mimic.mit.edu/docs/iv/modules/hosp/labevents/): 来自患者样本的实验室测量。

labevents 表存储为单个患者进行的所有实验室测量的结果。这些包括血液学测量、血气、化学面板和不太常见的测试，如基因检测。

hadm_id 是使用转院表分配给靠近住院的实验室的。然而，这并不总是能完美地捕捉到与住院时间相近的实验室。具体来说，从v2.1版开始，通过连接入院病人的 subject_id、入院时间和 disctime，可以为59,327,830个观测值分配 hadm_id。然而，这些观测值中只有58,131,956个（98%）有一个 hadm_id 实际存储在labevents表中。希望确保捕获与住院时间相近的实验室的用户应该意识到这一点，并在必要时使用带有时间的连接。

- charttime: 绘制实验室测量值的时间。这通常是获取标本的时间，通常明显早于可进行测量的时间。
- storetime: 在实验室系统中提供测量值的时间。这是信息提供给护理提供者的时间。
- itemid: 一个标识符，唯一地表示实验室概念
- value & valuenum: 实验室测量的结果，如果是数字，则转换为数字数据类型的值。

Links to

- d_labitems(另外一张表) on itemid

用药表

Out[17]:

	subject_id	hadm_id	starttime	stoptime	medication	frequency
0	15346117	24544765	13/1/2201 00:00:00	18/1/2201 14:00:00	Sodium Chloride 0.9% Flush	Q8H
1	15346117	24544765	13/1/2201 00:00:00	21/1/2201 19:00:00	Heparin	TID
2	15346117	24544765	13/1/2201 01:00:00	13/1/2201 16:00:00	Acetaminophen	Q6H:PRN
3	15346117	24544765	13/1/2201 01:00:00	21/1/2201 19:00:00	Amlodipine	DAILY
4	15346117	24544765	13/1/2201 01:00:00	21/1/2201 19:00:00	Aspirin	DAILY
...
667601	17614063	26391110	22/4/2177 14:00:00	22/4/2177 20:00:00	Fluzone Trival	ONCE
667602	12361593	28465050	16/7/2147 13:00:00	16/7/2147 19:00:00	Fluzone Trival	ONCE
667603	17050261	28079544	20/4/2184 15:00:00	20/4/2184 20:00:00	Fluzone Trival	ONCE
667604	18749775	26585465	22/7/2147 04:00:00	24/7/2147 12:00:00	COQ10 (300 mg)	Q 12H
667605	16112663	24076527	8/12/2167 08:00:00	8/12/2167 17:00:00	Tenecteplase	2X

667606 rows × 6 columns

pharmacy表条目:667606

pharmacy表统计subject数量: 12352

用药表中字段解释

pharmacy 更多指标解释 (https://mimic.mit.edu/docs/iv/modules/hosp/pharmacy/): 处方药的处方、剂量和其他信息。

药房表提供了有关开给患者的填充药物的详细信息。药房信息包括药物剂量、处方剂量数、给药频率、用药途径和处方持续时间。

- subject_id : 是一个唯一的标识符，它指定了一个单独的病人。任何与单个subject_id相关的行都与同一个人有关。
- hadm_id : 是一个整数标识符，对每个病人的住院治疗是唯一的。
- starttime stoptime : 给定的处方药的开始和停止时间
- medication : 提供的药物名称。
- frequency : 应向患者给药的频率。许多常用的短手都用在频率栏中。 Q#表示每隔#小时；例如“Q6”或“Q6H”是每 6 小时一班。

Links to

- poe on poe_id
- prescriptions on pharmacy_id
- emar on pharmacy_id

坍塌[Collapse]用药表

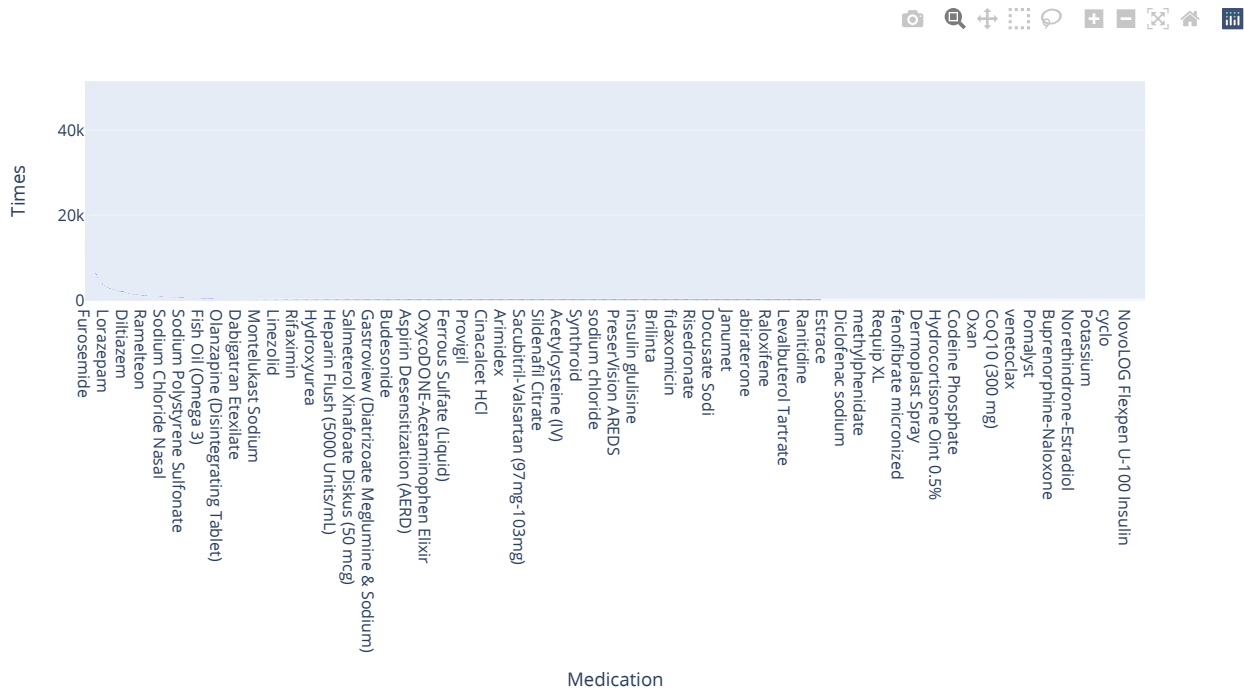
- 每个主体subject多条用药记录collapse一个

Medication的分布

Out[19]:

	Medication	Times
0	Furosemide	48905
1	Insulin	37613
2	Potassium Chloride	34367
3	Heparin	28422
4	Metoprolol Tartrate	26800
...
2010	Melphalan	1
2011	Critic-Aid Clear	1
2012	Benicar HCT	1
2013	Nasonex	1
2014	COQ10 (300 mg)	1

2015 rows × 2 columns



合并Medication&Frequency&Time

D:\Dev\anaconda3\envs\pytorch\lib\site-packages\pandas\core\groupby\generic.py:303: FutureWarning:

Dropping invalid columns in SeriesGroupBy.agg is deprecated. In a future version, a TypeError will be raised. Before calling .agg, select only columns which should be valid for the aggregating function.

Out[21]:

medication-frequency-time	
subject_id	
10000980	Sodium Chloride 0.9% Flush-Q8H:PRN-3/1/2188 1...
10001492	Influenza Virus Vaccine-ASDIR-23/9/2136 23:00:...
10001877	Influenza Virus Vaccine-NOW X1-22/11/2150 03:0...
10002013	Aspirin-ONCE-10/7/2160 21:00:00-10/7/2160 22:0...
10002155	nan-ASDIR-4/8/2129 14:00:00-4/8/2129 13:00:00,...
...	...
19996783	nan-ASDIR-9/5/2188 22:00:00-10/5/2188 03:00:00...
19997367	Sodium Chloride 0.9% Flush-Q8H-24/5/2127 21:0...
19997473	Sodium Chloride 0.9% Flush-Q8H-11/9/2173 03:0...
19998330	Sodium Chloride 0.9% Flush-Q8H:PRN-20/9/2178 ...
19998497	Ondansetron-ONCE-2/7/2139 21:00:00-3/7/2139 20...

12352 rows × 1 columns

展开 medication-frequency-time,

Out[22]: 'Sodium Chloride 0.9% Flush-Q8H:PRN-3/1/2188 19:00:00-5/1/2188 22:00:00,HydrALazine-ONCE-3/1/2188 20:00:00-4/1/2188 19:00:00,Clonidine Patch 0.1 mg/24 hr-QSUN-3/1/2188 10:00:00-3/1/2188 19:00:00,Clonidine-BID-3/1/2188 20:00:00-4/1/2188 11:00:00,Trandolapril-DAILY-3/1/2188 10:00:00-3/1/2188 19:00:00,Nifedipine CR-DAILY-3/1/2188 10:00:00-3/1/2188 19:00:00,HydrALazine-ONCE-3/1/2188 20:00:00-4/1/2188 19:00:00,HydrALazine-ONCE-3/1/2188 20:00:00-4/1/2188 19:00:00,Albuterol 0.083% Neb Soln-Q2H:PRN-3/1/2188 21:00:00-5/1/2188 22:00:00,Atenolol-DAILY-3/1/2188 10:00:00-3/1/2188 20:00:00,Clopidogrel-DAILY-3/1/2188 10:00:00-5/1/2188 22:00:00,Hydrochlorothiazide-DAILY-3/1/2188 10:00:00-3/1/2188 20:00:00,Nitroglycerin SL-PRN-3/1/2188 21:00:00-5/1/2188 22:00:00,Ranitidine-DAILY-3/1/2188 10:00:00-5/1/2188 22:00:00,Simvastatin-DAILY-3/1/2188 10:00:00-3/1/2188 21:00:00,Aspirin-DAILY-3/1/2188 10:00:00-4/1/2188 10:00:00,Multivitamins-DAILY-3/1/2188 10:00:00-5/1/2188 22:00:00,Dextrose 50%-PRN-3/1/2188 21:00:00-5/1/2188 22:00:00,Glucagon-Q15MIN:PRN-3/1/2188 21:00:00-5/1/2188 22:00:00,Insulin-ASDIR-3/1/2188 21:00:00-5/1/2188 22:00:00,Insulin-DINNER-3/1/2188 17:00:00-5/1/2188 22:00:00,Hydrochlorothiazide-DAILY-3/1/2188 10:00:00-3/1/2188 21:00:00,Metoprolol Tartrate-BID-3/1/2188 20:00:00-4/1/2188 08:00:00,Heparin-ONCE-3/1/2188 21:00:00-4/1/2188 10:00:00,Heparin-ASDIR-3/1/2188 21:00:00-4/1/2188 10:00:00,Heparin-ASDIR-3/1/2188 21:00:00-4/1/2188 10:00:00,Docusate Sodium-BID:PRN-3/1/2188 22:00:00-5/1/2188 22:00:00,Senna-BID:PRN-3/1/2188 22:00:00-5/1/2188 22:00:00,Atorvastatin-DAILY-3/1/2188 10:00:00-5/1/2188 22:00:00,Furosemide-ONCE-4/1/2188 01:00:00-5/1/2188 00:00:00,traZODONE-ONCE-4/1/2188 01:00:00-5/1/2188 00:00:00,Furosemide-ONCE-4/1/2188 09:00:00-5/1/2188 08:00:00,Metoprolol Tartrate-BID-4/1/2188 20:00:00-4/1/2188 17:00:00,Aspirin-DAILY-5/1/2188 10:00:00-5/1/2188 22:00:00,Lisinopril-DAILY-4/1/2188 10:00:00-4/1/2188 16:00:00,Heparin-TID-4/1/2188 14:00:00-5/1/2188 22:00:00,Furosemide-ONCE-4/1/2188 13:00:00-5/1/2188 12:00:00,Lisinopril-ONCE-4/1/2188 17:00:00-5/1/2188 12:00:00,Lisinopril-DAILY-5/1/2188 10:00:00-5/1/2188 22:00:00,Clonidine-BID-4/1/2188 20:00:00-4/1/2188 17:00:00,Nifedipine CR-ONCE-4/1/2188 17:00:00-5/1/2188 12:00:00,Nifedipine CR-DAILY-5/1/2188 10:00:00-5/1/2188 22:00:00,Metoprolol Tartrate-BID-4/1/2188 20:00:00-5/1/2188 12:00:00,Magnesium Sulfate-ONCE-5/1/2188 10:00:00-5/1/2188 22:00:00,Furosemide-DAILY-5/1/2188 10:00:00-5/1/2188 22:00:00'

ICU停留 表

Out[23]:

	subject_id	hadm_id	stay_id	first_careunit		last_careunit	intime	outtime	los
0	11679839	29636865	30021881	Coronary Care Unit (CCU)		Coronary Care Unit (CCU)	19/6/2112 14:41:32	20/6/2112 21:50:49	1.298113
1	13787728	27446327	30024343	Coronary Care Unit (CCU)		Coronary Care Unit (CCU)	17/2/2123 18:52:51	18/2/2123 13:39:32	0.782419
2	19352034	21509557	30058866	Coronary Care Unit (CCU)		Coronary Care Unit (CCU)	18/4/2114 02:02:38	30/4/2114 21:41:14	12.818472
3	18817948	23379351	30061662	Coronary Care Unit (CCU)		Coronary Care Unit (CCU)	2/11/2142 22:18:49	4/11/2142 00:46:50	1.102789
4	14964123	25103078	30081334	Surgical Intensive Care Unit (SICU)	Surgical Intensive Care Unit (SICU)		15/6/2137 16:33:00	17/6/2137 18:28:59	2.080544
...
4821	19420059	22049257	39920422	Cardiac Vascular Intensive Care Unit (CVICU)	Cardiac Vascular Intensive Care Unit (CVICU)		17/1/2114 08:55:42	22/1/2114 13:16:30	5.181111
4822	12673141	21284928	39921830	Cardiac Vascular Intensive Care Unit (CVICU)	Coronary Care Unit (CCU)		19/4/2131 05:28:27	23/4/2131 21:17:05	4.658773
4823	10836444	25551438	39934059	Cardiac Vascular Intensive Care Unit (CVICU)	Cardiac Vascular Intensive Care Unit (CVICU)		9/12/2170 13:50:58	10/12/2170 17:23:26	1.147546
4824	17014465	25960753	39966638	Cardiac Vascular Intensive Care Unit (CVICU)	Cardiac Vascular Intensive Care Unit (CVICU)		5/2/2112 09:23:57	6/2/2112 17:00:33	1.317083
4825	12275003	22562812	39992247	Cardiac Vascular Intensive Care Unit (CVICU)	Cardiac Vascular Intensive Care Unit (CVICU)		15/8/2182 09:37:33	16/8/2182 17:25:44	1.325127
4826 rows × 8 columns									

4.1 ICU停留表中字段解释

ICU stays 更多 (<https://mimic.mit.edu/docs/iv/modules/icu/icustays/>): ICU 停留的跟踪信息，包括入院和出院时间。

- stay_id 是一个生成的标识符，不是基于任何原始数据标识符。医院和ICU的数据库没有内在的联系，所以没有任何ICU遭遇标识符的概念。
 - icustays 表是由 transfer 表衍生出来的。具体来说，它排除了病房不是ICU的转院记录。
 - transfer 表中多个连续的ICU住院条目被合并为 transfer 表中的一个条目。
 - 心率测量被用来确定重症监护室住院是否有效。如果 transfer 表中的重症监护室停留的心率测量值不可用，重症监护室停留将不包括在 icustays 表中 (~5%) 。
-
- subject_id & hadm_id: 指定患者的标识符: subject_id 对于患者是唯一的， hadm_id 对于患者住院是唯一的， stay_id 对于患者住院是唯一的。
 - first_careunit & last_careunit: FIRST_CAREUNIT 和 LAST_CAREUNIT 分别包含病人被护理的第一个和最后一个ICU类型。由于一个stay_id将所有的在24小时内入院的重症监护室分组，所以病人有可能从一种重症监护室转到另一种重症监护室而拥有相同的 stay_id 。
 - intime & outtime: INTIME 提供患者被转移到 ICU 的日期和时间。 OUTTIME 提供患者转出 ICU 的日期和时间。
 - los: LOS 是患者在给定的 ICU 住院期间的住院时间，可能包括一个或多个 ICU 病房。停留时间以小数天数计算。

Links to:

- patients on subject_id
- admissions on hadm_id

```
D:\Dev\anaconda3\envs\pytorch\lib\site-packages\pandas\core\groupby\generic.py:303: FutureWarning:
Dropping invalid columns in SeriesGroupBy.agg is deprecated. In a future version, a TypeError will be raised. Before calling
.agg, select only columns which should be valid for the aggregating function.
```

Out[24]:

	subject_id	los	icustay_inout_time
0	10002155	6.178912	4/8/2129 12:45:00-10/8/2129 17:02:38
1	10002495	5.087512	22/5/2141 20:18:01-27/5/2141 22:24:02
2	10005817	2.359097	15/12/2132 09:29:01-17/12/2132 18:06:07
3	10007058	2.022963	7/11/2167 20:22:00-9/11/2167 20:55:04
4	10009686	1.446181	30/4/2164 07:37:10-1/5/2164 18:19:40
...
4292	19994379	2.929421	9/5/2131 00:05:08-11/5/2131 22:23:30
4293	19995780	3.383530	20/10/2125 11:32:13-23/10/2125 20:44:30
4294	19996783	3.225023	10/5/2188 02:20:42-10/5/2188 22:30:28,11/5/218...
4295	19997473	8.290440	17/9/2173 15:07:55-22/9/2173 06:39:09,13/9/217...
4296	19998330	0.851690	20/9/2178 21:40:00-21/9/2178 18:06:26

4297 rows × 3 columns

Chart和Lab表中的itemid关联Label

任务：由于Chart表&Lab表中的itemid没有具体的解释，需要分别关联两张表itemID&Labels来获取具体的解释，作者姑且认为是检查项目；讨论的过程得知，Chart表是进ICU的检查(此观点仍需进一步确认).

特性：展示了Chart&Lab两张表的Label具体的分布情况（以做过该项检查病人数量为Y轴数据），同时展示了前10个检查项目

所有的Chart itemid类别1889

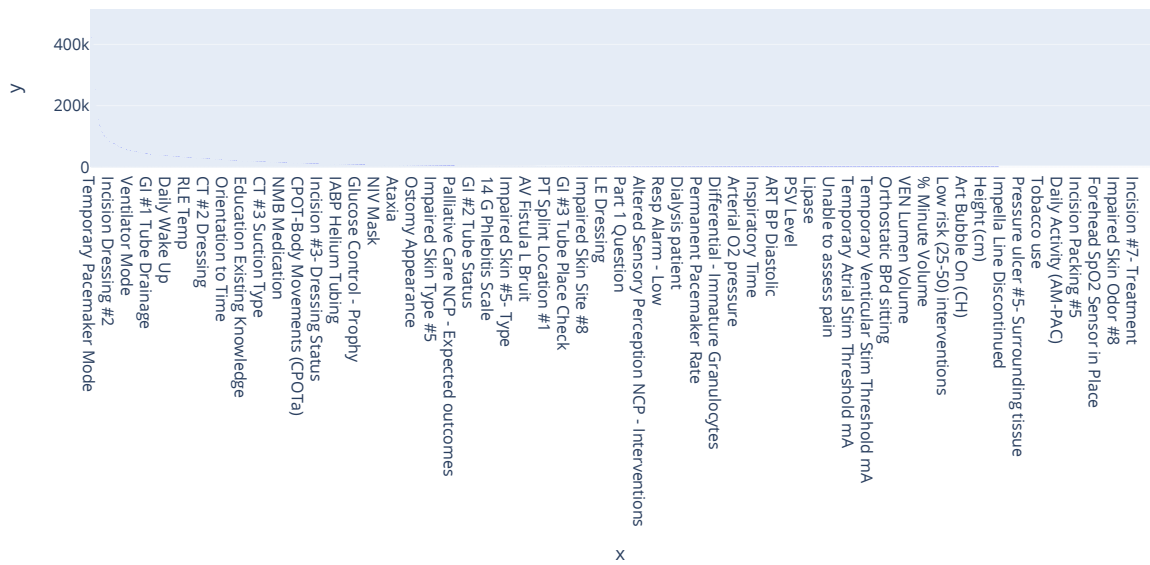
Out[25]: True

Out[26]:

	subject_id	hadm_id	charttime	storetime	itemid	value	valuenum	label
5588786	14872672	29294271	27/11/2154 18:00:00	27/11/2154 17:11:00	227948	Soft limb	NaN	Restraint Device (Non-violent)
9016370	16011666	25016707	26/7/2134 09:00:00	26/7/2134 09:10:00	220210	18.0	18.0	Respiratory Rate
6957243	17348907	23317599	1/8/2176 15:01:00	1/8/2176 15:05:00	220179	152.0	152.0	Non Invasive Blood Pressure systolic
1512429	10219508	24008925	17/12/2150 13:00:00	17/12/2150 13:21:00	223761	97.2	97.2	Temperature Fahrenheit
4496182	17288844	25595502	30/8/2197 08:17:00	30/8/2197 09:17:00	227367	1.0	1.0	18 Gauge Dressing Occlusive
12185427	15855896	24897676	20/11/2174 12:00:00	20/11/2174 11:59:00	228496	Lower leg - Left	NaN	Impaired Skin #1- Location
17768708	11417242	29307193	16/2/2178 10:23:00	16/2/2178 10:23:00	227292	1.0	1.0	Arterial Line Dressing Occlusive
8408073	18896613	21855159	17/4/2175 15:11:00	17/4/2175 15:11:00	228096	-1 Awakens to voice (eye opening/contact) > 10...	-1.0	Richmond-RAS Scale
13475135	18098128	25408525	21/5/2141 10:34:00	21/5/2141 10:34:00	225185	1.0	1.0	Skin Care
12020170	14527409	24099405	22/7/2203 12:00:00	22/7/2203 12:06:00	227957	No	NaN	Signs of Injury from Intervention

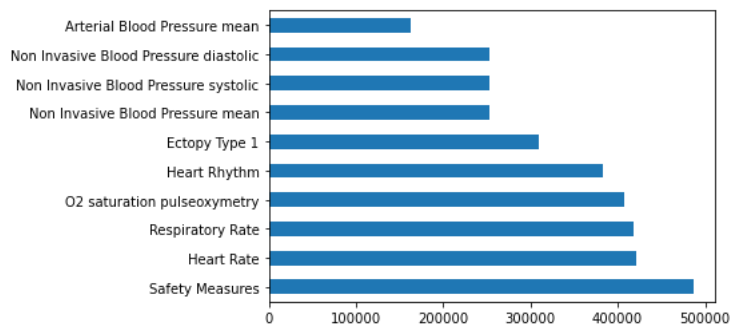
总计ICU的Label种：1842

Chart-Label 分布情况统计



前10项ICU检查

Out[29]: <AxesSubplot:>



所有的实验室itemid类别655

Out[30]: True

Out[31]:

	subject_id	hadm_id	charttime	storetime	itemid	value	valuenum	label
320230	19393286	27673374	26/8/2132 06:00:00	26/8/2132 07:10:00	50861	14	14.0	Alanine Aminotransferase (ALT)
1153347	11973413	29660298	26/11/2167 07:35:00	26/11/2167 07:50:00	51277	13.6	13.6	RDW
1920802	14645012	29765581	23/12/2177 17:45:00	23/12/2177 17:47:00	50804	24	24.0	Calculated Total CO2
1539038	13408838	20225357	7/6/2127 08:00:00	7/6/2127 11:50:00	50911	6	6.0	Creatine Kinase, MB Isoenzyme
1972708	16162344	26252203	20/10/2125 17:05:00	20/10/2125 17:10:00	51249	35.2	35.2	MCHC
1723738	18860372	25940510	14/12/2120 04:00:00	14/12/2120 04:45:00	50947	1	1.0	I
790400	11737174	27174306	11/2/2147 08:40:00	11/2/2147 09:18:00	51275	47.9	47.9	PTT
2230090	13559413	21528106	17/12/2163 08:30:00	17/12/2163 09:14:00	51277	15.7	15.7	RDW
2673388	15784822	28147695	6/11/2116 04:45:00	6/11/2116 05:45:00	51237	1.8	1.8	INR(PT)
2824295	11205212	25178486	29/7/2174 15:44:00	29/7/2174 22:39:00	51706	NaN	NaN	Problem Specimen

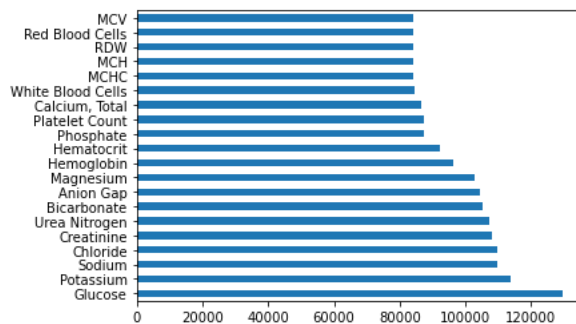
总计label(检查项目)种: 571

实验-Label 分布情况统计



前10项实验室检查

Out[34]: <AxesSubplot:>



MISC: 合并ICU检查和实验室检查表

观察&思考

1. Chart表的数据条目比实验室Lab表要多 15million (15,307,110); 实验室条目: 3,197,442; chart条目: 18,504,552
2. Lab表中的病人为: 12,257; Chart表中病人为: 4,297
3. Chart表中的病人都能在Lab表中找到吗?

- 计算结果是 4288, ICU病人总计4,297; 有9个chart病人有chart检查没有Lab检查
- 4. 实验室中的itemid有655中, Chart中有1889种, 仍然符合chart多, 进ICU少 (lab) .

推测: chart检查的项目多,进ICU的病人少, 但是检查更多

5. ItemId以 5**** 开头的是lab表, 表示实验室检查; itemId以 2***** 开头为Chart表, 表示ICU检查

- 检查'5'开头的itemId有多少, 是否和Lab表的655同; 同理检查Chart表 (ICU,从上文得知ICU) 【检验, 正确】

6. ICU检查和实验室检查会重叠吗?

- 计算的结果说明这两个表的itemId没有发现重叠, 结果为0.

运算: $655 \times 1889 = 1237295$

- 说明总共的检查项目是 $655 + 1889$

7. 一个病人做的多项检查也如同 info4表中的长数据转变成宽数据吗, 也就是label值?

labels值有很长，从实验室到ICU加起来(655+1889).同一个人的实验室检查合成一条记录，ICU检查也合并成一条数据；

8. 以itemid为基础进行合并吗？合并完成后以 subject_id 排序，不同的检查对应了一个single patient.

最后采取的操作是: '连接': 理由是itemid可以全部区分开；两个表中的itemid没有交集

合并之后的检查项目表

Out[37]:

	subject_id	hadm_id	charttime	storetime	itemid	value	valuenum	label
20270947	10000980	29654838	4/1/2188 03:56:00	4/1/2188 05:25:00	50970	4.9	4.900	Phosphate
20270957	10000980	29654838	5/1/2188 06:09:00	5/1/2188 08:29:00	50912	1.4	1.400	Creatinine
20270958	10000980	29654838	5/1/2188 06:09:00	5/1/2188 08:29:00	50960	1.8	1.800	Magnesium
20270959	10000980	29654838	5/1/2188 06:09:00	5/1/2188 08:29:00	50970	5.3	5.300	Phosphate
20270960	10000980	29654838	5/1/2188 06:09:00	5/1/2188 08:29:00	50971	4.3	4.300	Potassium
...
19581388	19998497	29288061	2/7/2139 17:26:00	2/7/2139 21:17:00	51492	30	30.000	Protein
19581389	19998497	29288061	2/7/2139 17:26:00	2/7/2139 21:17:00	51493	2	2.000	RBC
19581390	19998497	29288061	2/7/2139 17:26:00	2/7/2139 21:17:00	51498	1.026	1.026	Specific Gravity
19581414	19998497	29288061	4/7/2139 02:40:00	4/7/2139 04:12:00	50960	2.2	2.200	Magnesium
19581387	19998497	29288061	2/7/2139 17:26:00	2/7/2139 21:17:00	51491	6.5	6.500	pH

21701994 rows × 8 columns

处理: 用药表&ICU停留表

- 1. 检查项目表 (Chart/Lab): uni_itmes_label (合并chart和lab)
- 2. 基本信息中的病人数量: 12428; 检查项目表的病人数量是 12,266; 用药表中的病人数量: 12,352; ICU停留表中的病人有: 4297例
- 3. 检查项目表表中的病人能够在用药表中找到的人数为 12,248
- 4. 用药表中病人号可以在基本信息中的找到的有: 12,352
- 5. 检查项目表的病人号可以在基本信息中的找到的有:
- 6. ICU的条目是4826

完整表