**Sepsis 1.0和Sepsis 3.0两种诊断标准与脓毒症病人死亡率的关系**

符合sepsis1.0的研究对象**纳入标准**：病人需要同时满足感染和SIRS标准大于或等于两条。即sepsis1.0 = 感染 + SIRS大于等于两条。

感染的条件：“感染（完整版）”excel表格里面的所有ICD编码，共832个。

SIRS标准（入住ICU第一个24h的数据）：在上述满足感染条件的病人中筛选出符合①体温＞38℃或＜36℃；②心率＞90次/分；③呼吸＞20次/分或二氧化碳分压PaCO2＜4.3kPa（32mmHg）；④白细胞数WBC＞12.0×10.9／L或＜4.0×10.9/L或幼稚细胞＞10%。符合两条及两条以上者可进入纳入标准。

注意事项：因为SIRS标准里的数据需要的是入住ICU第一个24h内的数据，所以上述数据的采集时间应该与病人入住ICU的时间对比，提取出第一个24h内的数据，可能会有多次测量，所以应该采用最符合的那个指标（也就是最差的那个指标），看其是否符合SIRS标准的要求。如，24h的心率值分别是70，78，99，100等，应该拿100这个值来与SIRS标准进行对比。另外，PaCO2的单位有kPa和mmHg两种，查询时需要注意区分。

符合sepsis3.0的**纳入标准**：病人需要同时满足感染和SOFA大于等于2分。

即sepsis3.0 = 感染 + SOFA大于等于两分。

感染的条件：同上。（即“感染（完整版）”excel表格里面的所有ICD编码，共832个。）

SOFA评分（入住ICU第一个24h）：SOFA大于等于2（注意 : SOFA也需要入住ICU第一个24和内的数据，所以SOFA的测量时间也应该与病人ICU入住时间进行对比，找出第一个24h的数据）

在MIMICIII数据库中分别筛选出符合两种诊断标准的所有成年病人。大于18岁

**排除标准**：我们只需要成人数据，所以排除入住ICU年龄在18岁以下的病人。（这个条件是从一开始就加以限定还是对符合上述标准的病人进行排除，由数据提取人员视情况决定）。

排除数据不全的病人（应该根据提取出的数据项具体情况决定），数据不全的病人是直接剔除还是用均值代换？（待定，将根据数据不全的病人所占的比例决定。）

纳入标准的具体信息：见excel表，请参照上述纳入标准进行数据挖掘。

满足纳入标准的研究对象，我们需要的主要数据：

　　病人ID（SUBJECT\_ID，HADM\_ID和ICUSTAY\_ID），入住ICU时的年龄，性别，种族，ICU类型，入住ICU时间，离开ICU时间， 死亡时间，ICU停留时间，SOFA具体分值，APACHE-II评分值，APACHE-III评分值，SAPSI评分值。（具体信息见excel表）

数据说明：1.我们需要用病人死亡时间与其入住ICU时间和离开ICU时间比较，得出所有研究对象的死亡率（离开ICU时的死亡率，21天死亡率和最终死亡率）。

2.几个评分值是用来对我们的统计模型进行校正和描述符合纳入标准的研究对象的基本特征的。

3.其余数据是用来描述合纳入标准的研究对象的基本特征的。

注意事项：

1.采集时间是指CHARTEVENTS表格中的CHARTTIME，我们需要的是与病人入住ICU时间（ICUSTAYS里的INTIME）对比不超过２４ｈ的，符合上述SIRS标准和SOFA标准的病人。

2. 实验室检查结果：实验室检查如动脉血二氧化碳分压和白细胞计数，CHARTEVENTS和LABEVENTS表格中都有，也就是说有ITEMID和LABITEMID两种项目ID，至于哪个表格里面的数据更全面，需要查询后再判断。另外，幼稚细胞数的结果，只有LABITEMID，所以可能只有LABEVENTS表格中有。（建议先找几个病人，然后分别查询一下两个表格中的实验室检查数据，哪个更多就采用哪个表格的数据，如白细胞计数在CHARTEVENTS中24h内有3个数据，LABEVENTS表格中有5个，那就用LABEVENTS的数据进行病人筛选。）

3. 关于ITEM＿ID的问题：尚不确定用SQL查询数据时，ITEMID是否要用，或者说是否全用(有的可能可以直接查询)，如果用的话，有的数据项可能不止一个ITEM\_ID（相同的项目 ，CV监护系统和MV监护系统有不同的ID，另外还有英文大小写及表达方法的问题）,查询时应该尽量全面。

4. ICD-9编码的问题:所有与感染相关的ICD-9编码已经在“感染（完整版）”excel表格里面列出，共832个，不排除出现某些疾病代码检索不出来病人的情况，因为有的疾病诊断在临床上罕见或者不常用，但是必须一一筛选。

5.为了避免错误的发生，建议由我和负责数据挖掘的人员分别进行相关数据ITEMID和的查询，然后再对比，这样可以减少个人失误。（在查询ITEMID时，如果查到那一行数据里LINKSTO是空的，也就是说这个ITEMID并没有在其他表格中用到，那这个ITEMID就是无效的）

6. 对于SOFA，APACHE和SAPSI评分系统，虽然前两者有对应的项目ID，但是只有MV监护仪的有，CV监护仪的没有，参考文献中对MIMIC－II评分分值的查询，是没有用项目ID，直接查询提取的，MIMIC III尚不清楚，查询时请注意该问题。（猜测可能是MIMIC II用的CV监护仪，没有项目ID，是进行直接查询的，MIMIC　III用的MV监护仪，是新编的项目ID，所以这部分数据的具体查询方法需要进一步确定，是分两部分查询还是可以统一查询。）

7.因为上述几个评分系统都是ICU常用的评分，所以应该大部分的病人都是具有这些数据的，可以利用这一点对上一条注意事项中的数据查询方法是否正确进行验证。

8.MIMIC数据库中时间信息都是被转换的。（具体转换机制可以参考Data Descriptor:MIMIC-III, a freely accessible critical care database这篇文献中Deidentification的部分）。

9.因为科研需要，在数据提取的过程中，原始的提取步骤及提取编码请尽量保存，需要保存的信息还包括每个步骤筛选出了多少病人，排除掉了多少病人，排除的原因等等，数据提取过程信息越完整越好。