

ICS 11.040.99
CCS C30

团 体 标 准

T/BISSC 001—2025

感染性休克全周期队列建设标准

Construction standard for the entire life-cycle cohort of septic shock

2025 - 09 - 30 发布

2026 -01- 01 实施

四川省生物信息学学会 发 布

目 次

前 言	III
引 言	IV
1 范围	5
2 规范性引用文件	5
3 术语和定义	5
3.1 脓毒症 sepsis	5
3.2 感染性休克 septic shock	5
3.3 队列 cohort	6
3.4 队列研究 cohort study	6
3.5 SOFA 评分 SOFA score	6
3.6 qSOFA 评分 qSOFA score	6
3.7 数据 data	6
3.8 数据集 dataset	6
4 队列构建标准	6
4.1 建设目标	7
4.2 队列设计	7
4.3 队列设置	9
4.4 研究对象	10
4.5 数据采集体系	11
4.6 质量控制	12
4.7 伦理审查和人类遗传资源管理	14
5 数据要素构成	17
5.1 电子病历数据	17
5.2 医学影像	18
5.3 组学数据	21
5.4 数据标准化存储	21
6 数据汇集	22
6.1 数据采集设备	22
6.2 数据采集过程	23
7 数据治理	24
7.1 数据脱敏	24
7.2 数据处理	27
7.3 数据管理	29
7.4 数据标注	30
8 数据计算	33
8.1 数据计算引擎	33
8.2 数据计算流程	34
8.3 确定数据计算目标	35
8.4 数据计算前处理	35
8.5 数据计算实施	36
8.6 计算结果呈现	37

9 互操作性	38
9.1 数据脱敏互操作	38
9.2 数据处理互操作	38
9.3 数据计算互操作	38
10 数据安全	39
10.1 数据访问权限控制	39
10.2 数据加密	39
10.3 安全审计	40
10.4 数据操作日志	40
10.5 数据备份	40
11 数据共享	41
附录 A	42
A.1 基本信息	42
A.2 现病史	42
A.3 既往史	42
A.4 检验检查结果	43
A.5 治疗情况	47
A.6 疗效评价	48

前　　言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由四川省生物信息学学会提出。

本文件由四川省生物信息学学会知识产权与标准化工作委员会归口。

本文件起草单位：四川大学华西医院、四川大学、上海交通大学医学院附属仁济医院、浙江大学医学院附属第二医院、山东省公共卫生临床中心、哈尔滨医科大学附属第一医院、河南省人民医院、中国科学院计算技术研究所、中南大学。

本文件主要起草人：康焰、曾筱茜、雷印杰、皋源、黄曼、宋璇、赵鸣雁、秦秉玉、赵屹、黄佳琦、刘忠禹、王俊人。

引　　言

感染性休克（septic shock）亦称脓毒性休克，是脓毒症的一个亚型，即感染引起机体反应失调导致的多器官功能障碍伴有循环衰竭和细胞代谢异常。据最新统计，全球脓毒症年发病4890万，死亡1100万，死亡率约22%，而感染性休克死亡率则可达30-50%。因此，针对感染性休克构建全周期大规模队列、开展相关研究具有重要的临床价值和社会价值。

本数据规范以《2021国际脓毒症与脓毒性休克管理指南》作为主要参考依据，旨在形成共识性的感染性休克全周期队列建设标准体系，为感染性休克的大规模队列构建、数据要素构成、数据的汇集、治理、计算及安全等提供一致性的行业标准，实现同质化的数据储存和使用，形成高质量的感染性休克标准数据集来支撑包括临床医学研究及人工智能医疗器械等研发和应用的需求。

感染性休克全周期队列建设标准

1 范围

本标准主要规定了感染性休克全周期队列和数据建设的基本要素和相关规范。

本标准涉及的数据类型包括电子病历、医学影像、人口社会学数据、随访数据等，涉及临床遗传变异、转录组、代谢组、单细胞转录组等组学数据及其他类型数据暂未纳入本标准的数据的数据规范。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

T/CPMA 003—2019 大型人群队列现场调查管理技术规范

T/CPMA 004—2019 大型人群队列终点事件长期随访技术规范

郭彧,李立明.大型人群队列研究技术规范[M].人民卫生出版社,2019

Elm E v, Altman D G, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007; 335: 806.

T/CESA 1109—2020 智能医疗影像辅助诊断系统技术要求和测试评价方法

T/EGAG 016—2022 智慧医疗影像辅助诊断平台质量评估规范

T/CMDA 001—2020 肝胆疾病标准数据规范：肝癌CT/MRI影像采集和处理标准

T/CMDA 002—2020 肝胆疾病标准数据规范：肝癌CT/MRI影像标注和质控标准

T/BISSC 01—2022 专科疾病标准数据集建设规范

T/CMEAS 045-2025 脓毒症多组学数据库建设及应用

GB/T 39725-2020 信息安全技术 健康医疗数据安全指南

3 术语和定义

以下术语和定义适用于本文件。

3.1 脓毒症 sepsis

因感染引起的宿主反应失调导致的危及生命的器官功能障碍。

3.2 感染性休克 septic shock

脓毒症经充分的液体复苏，仍需使用血管活性药物以维持平均动脉压 $\geq 65\text{mmHg}$ ，且血乳酸浓度 $> 2\text{mmol/L}$ 。

注：感染性休克是脓毒症引起的严重循环和细胞代谢紊乱综合征，又称脓毒性休克，分类属于分布性休克。

3.3 队列 cohort

流行病学研究中，需要进行随访的某一特定人群。

3.4 队列研究 cohort study

将某一特定人群按是否暴露于某可疑因素或暴露程度分为不同的亚组，追踪观察两组或多组成员结局（如疾病）发生的情况，比较各组之间结局发生率的差异，从而判定这些因素与该结局之间有无因果关联及关联程度的一种观察性研究方法。

3.5 SOFA 评分 SOFA score

全称为序贯器官衰竭评估（Sequential Organ Failure Assessment）评分，是用于定量和客观描述重症患者随时间推移发生器官功能障碍或器官衰竭程度的量表，可用于脓毒症患者筛查。

[Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 1996;22(7):707-710.]

3.6 qSOFA 评分 qSOFA score

全称为快速序贯器官衰竭评估（quick Sequential Organ Failure Assessment）评分，可用于疑似脓毒症患者的快速床旁筛查。

3.7 数据 data

信息的可再解释的形式化表示，以适用于通信、解释或处理。

[GB/T 25000.12/2017，定义4.2]

3.8 数据集 dataset

数据记录汇集的数据形式。

[GB/T35295-2017，定义2.1.46]

4 队列构建标准

感染性休克全周期队列建设应确保科学性、规范性、可行性和可推广性，本节提出感染性休克全周期队列建设框架，主要包括建设目标、队列设计、队列设置、研究对象、数据收集、质量控制及伦理审查和人类遗传资源管理七方面。

4.1 建设目标

感染性休克全周期队列建设要具备明确目标，需阐述研究对象、暴露因素、结局事件以及具体的研究假说或者研究问题。如果在队列建设的早期阶段，目标不明确，应至少明确研究对象、暴露因素、结局事件。感染性休克全周期队列建设目标是阐明“感染-脓毒症-感染性休克-转归”的疾病演进规律。具体而言，针对因感染入院就诊的患者（研究对象），广泛收集其全周期实时动态信息（暴露因素），以阐明感染性休克的疾病演进规律及后续转归情况（结局事件）。

4.2 队列设计

4.2.1 队列类型

在队列建设中，队列类型的选择对于研究的设计、实施和结果的解读至关重要。一般来说可以按照研究对象和数据收集的时间点来划分。按照队列参与人群划分，常见的队列类型有专病队列、自然人群队列和出生队列等。感染性休克全周期队列研究对象为因感染入院就诊的患者，因此其队列类型应为专病队列。如按照数据收集的时间点则可分为前瞻性队列、回顾性队列和双向性队列。感染性休克全周期队列可采用回顾性、前瞻性，或两者结合的双向性队列设计，研究者可根据自身的队列实施条件选择合适的队列类型。

4.2.1.1 回顾性队列

研究者利用已有的医疗记录、数据库等资料，根据研究开始时已经掌握的有关研究对象在过去某个时间点的暴露状况的历史材料进行回顾性分析，此时研究结局已经发生，不需要再进行前瞻性的随访观察。例如，感染性休克全周期队列研究采用回顾性队列设计，可利用全国多家医院的电子病历系统，回顾性收集过去五年内所有因感染入院就诊的患者的临床资料，将患者入院 24 小时内是否发展为脓毒症乃至感染性休克作为研究结局，分析影响感染患者向脓毒症、感染性休克转归的危险因素。回顾性队列可以快速、高效地利用已有的数据资源，进行大样本量的研究，但数据质量可能受到限制。

4.2.1.2 前瞻性队列

研究者根据预先设定的纳入和排除标准，前瞻性招募目前未发生研究结局的研究对象，收集基线资料，并按照一定的随访计划，定期收集相关数据，观察研究对象在随访期内目标疾病的发生情况。关键在于，当研究的暴露因素完成收集时，研究结局尚未发生，需要前瞻性观察一段时间才能获得。例如，感染性休克全周期队列研究采用前瞻性队列设计，可选取全国多家医院，纳入符合感染诊断标准，但尚

未发展为脓毒症和感染性休克的患者。在入组时，收集患者的详细信息，包括人口学资料、基础疾病、感染部位、病原学结果、SOFA 评分、qSOFA 评分等。随后，按照预先设定的随访计划（例如每 24 小时评估一次），动态监测患者的病情变化、治疗措施、器官功能支持情况等，直至患者出院或死亡。通过前瞻性随访，可以实时观察感染患者向脓毒症、感染性休克发展的动态过程，并分析影响其转归的因素。

4.2.1.3 双向性队列

研究者在回顾性队列研究的基础上，如果从暴露到现在的观察时间还不能满足研究的要求，如结局事件还没有发生或没有完全发生，还需继续前瞻性观察一段时间时，则选用双向性队列。双向性队列将前瞻性和回顾性队列研究结合起来，因此会具有上述两类的优点，而且相对的在一定程度上弥补了各自的不足。例如，感染性休克全周期队列研究采用双向性队列设计，可先选取全国多家医院，回顾性收集过去五年内所有因感染入院就诊的患者的临床资料，将患者入院 24 小时内是否发展为脓毒症乃至感染性休克作为研究结局，完成回顾性队列的构建，并进一步跟踪患者的 3 年内是否发生死亡，以补充回顾性队列研究中缺少纵向随访结局的缺陷。

4.2.2 研究对象

研究对象是目标人群中具有代表性的一部分（符合队列研究纳入条件和标准），通常根据是否暴露于某可疑因素或暴露程度被分为不同的亚组（如暴露组和非暴露组）。感染性休克全周期队列是旨在阐明“感染-脓毒症-感染性休克-转归”的疾病演进规律的专病队列，其研究对象为因感染入院就诊的患者。可基于该专病队列开展多个子队列研究，如纳入因感染入院的患者作为研究对象，并根据年龄分为不同的亚组，随访其是否发生脓毒血症，进行组间结果比较。

4.2.3 结局事件

结局事件指根据研究目标所设定的研究对象出现的预期结局，感染性休克全周期队列的结局事件包括主要结局、次要结局以及竞争风险事件三类。

4.2.3.1 主要结局

主要结局应该与队列研究目标对应，是研究中最关注的结局事件，也是整个研究最终要解决或回答的问题。在感染性休克全周期队列的研究中，如果以脓毒症或感染性休克疾病发生作为主要结局，则应该明确疾病的临床诊断标准，以确定结局事件的发生。

脓毒症的诊断标准应同时满足以下 2 条：

- a) 确诊感染或疑似感染；
- b) SOFA 评分较基线增加≥2 分。

感染性休克的诊断标准应同时满足以下 3 条：

- a) 脓毒症诊断成立；
- b) 充分液体复苏后仍需使用血管活性药物以维持平均动脉压 ≥ 65 mmHg；
- c) 血乳酸浓度 > 2 mmol/L。

4.2.3.2 次要结局

次要结局也与整个队列研究的目的相关联，通常紧跟主要结局出现，作为整个研究的支持性证据，常见的次要结局包括 30 天内再入院、死亡等。次要结局不宜太多，可以与主要结局事件相同但时间维度不同。比如在研究某种药物对感染性休克患者的治疗效果，可以用药后 30 天内是否死亡作为主要结局事件，并以用药后半年内是否死亡作为次要结局事件。

4.2.3.3 竞争风险事件

竞争风险事件是指在观测主要结局或次要结局事件是否发生的过程中，出现了阻碍观测结局发生的其他事件，即与研究结局形成了“竞争”关系。这类事件与结局具有相互排斥性，常见如死亡、其他严重疾病或接受其他治疗干预等。例如，在观测感染性休克的发生过程中，若患者因恶性肿瘤终末期死亡，则死亡事件阻断了感染性休克的观察，构成了竞争风险事件。需要注意的是，“失访”不应一概归为竞争风险事件。如因迁居、主动退出研究等随机因素导致的失访，仅造成信息缺失，并不阻断研究结局的实际发生，不属于竞争风险事件。而若失访由死亡、严重并发症、病情恶化等非随机因素导致，确实阻碍了研究结局的发生，这类情况应被视为真正的竞争风险事件。

4.3 队列设置

4.3.1 实施场所

为确保感染性休克全周期队列的建设质量和代表性，实施场所（参与的医疗机构）的遴选应遵循科学、全面和可行的原则，具体细则如下：

- a) 建议选择两家及以上的医疗机构作为实施场所以满足研究结果的外部有效性。例如，可选择华东、华北、华中、华南和西北等地区的代表性医院。此外，建议涵盖经济发达区和中低收入区域医院，确保队列样本能代表全国感染性休克患者实际情况；
- b) 优先选取三级甲等医院、综合性医院等具备重症医学科、感染科、急诊科等关键科室的医疗机构，确保医院具备收治和管理感染性休克患者的能力；
- c) 参与医院应有完善的重症救治体系和完善的医院感染管理制度，能够准确收集基线及随访数据，并满足多学科协作（如急诊、感染科、重症医学科）的要求；
- d) 参与医院应配备成熟的电子医疗记录系统及数据管理平台，确保患者数据的准确、完整和时效性；

e) 优先选择具有既往参与多中心队列研究或大型临床科研项目经验的医院，以保证现场执行能力和随访管理质量。

4.3.2 实施时间

感染性休克全周期队列设计需明确划分四个基本阶段：研究对象招募、基线调查、随访调查和结局评估。阶段划分有助于确保研究流程的系统性和可操作性，为高质量数据采集和分析提供时间保障。

4.3.2.1 研究对象招募

根据各参与医院的患者数量、人力资源、物力资源等情况，确定一个统一的研究对象招募时间窗口，应跨越至少一个完整的自然年（例如：2024年1月至2024年12月），以确保纳入队列的感染性休克患者能够覆盖不同地区、不同季节的特点和变化，具有更好的代表性。

4.3.2.2 基线调查

所有入组患者在完成招募后24小时内完成基线数据采集，以保证基线数据的有效性。

4.3.2.3 随访调查

随访计划应根据研究目标确定合理的追踪期，并明确随访的关键时间节点（如28天、90天、出院后6个月等），系统记录患者在“感染-脓毒症-感染性休克-转归”的全周期内病情变化、治疗效果与预后情况等信息，为纵向动态分析提供支持。

4.3.2.4 结局评估

应设置明确的结局评估时间点，统一收集和判定患者的主要和次要结局事件（如死亡率、器官功能恢复情况、再入院、生活质量等）。结局评估是队列研究终点，需确保资料完整性和评估一致性，以保障研究结论的科学性和有效性。

4.4 研究对象

4.4.1 入选标准

应能满足研究目的的需要，能保证长期跟踪随访。具体满足以下条件：

- a) 符合调查对象的纳入要求，详见4.2.3.1；
- b) 可获取研究对象的知情同意书。

4.4.2 招募登记

4.4.2.1 招募登记流程

招募登记流程应保证以下要素，具体内容及方式可根据客观条件进行适当调整：

- a) 对调查对象的积极参与表示欢迎和感谢;
- b) 介绍研究项目内容、所获得的收益及可能存在的风险, 获取其知情同意, 并签署知情同意书;
- c) 按照研究设定的纳入和排除标准核实调查对象;
- d) 收集足够的信息以保证后期随访工作的顺利开展;
- e) 使用专项软件或专用汇总表格登记其基本信息, 发放调查所需物品。

4.4.2.2 注意事项

招募登记环节的注意事项包括:

- a) 如实向调查对象介绍项目的内容、意义等重要信息, 不得隐瞒、欺骗;
- b) 填写表单时字迹应清晰完整;
- c) 签署的《知情同意书》应妥善保管;
- d) 对调查对象的个人信息严格保密。

4.5 数据采集体系

4.5.1 操作规范

4.5.1.1 基线调查

(1) 现场调查。在患者首次入院或确诊时, 采用统一标准的问卷调查、结构化面谈、标准量表以及临床体格检查等多种方式同步采集信息。现场调查应由经过专门培训的数据收集人员或护士执行, 确保所有数据按统一格式记录, 保证基线数据的统一性与可比性。

(2) 现场检查。现场检查包括常规体检(血压、体温、心率等)以及必要的实验室检测(例如血乳酸水平、炎症指标、器官功能指标)、影像学检查等, 确保每项检测均依据预先设定的标准执行, 并实时录入电子记录系统。

(3) 信息系统数据整合。通过与医院信息系统数据对接, 实现数据自动抓取、预处理和格式标准化, 保证信息的全面性与连续性。

4.5.1.2 随访调查

(1) 院内随访。前瞻性队列或双向性队列的前瞻阶段, 研究对象入组时在院内完成基线调查后, 后续随访主要在各参与医院的相关科室(如重症医学科、普通病房、急诊)进行。现场随访方式包括医生面访、量表调查、实验室检测以及影像学检查, 从而获取连续、动态的临床数据。

(2) 长期或院外随访。对于出院患者以及需要长期随访的患者, 可通过电话、短信、网络问卷、微信公众号、专用随访App等方式对患者进行远程随访, 收集结局事件、再入院、死亡、迁移和失访情况等信息。此外, 可借助社区卫生服务机构或当地卫生和医保部门的合作, 利用其定期门诊或体检机会、

行政管理电子数据库等进行补充随访，确保数据全面收集。其注意事项包括：

- a) 对认知障碍患者，必须获得法定监护人的书面同意并全程录音留存；
- b) 随访内容新增生物样本采集或数据共享（如社区医疗记录），需重新签署补充知情同意书；
- c) 电话/视频随访需在安静环境进行，明确告知通话内容将被录音；
- d) 网络问卷设置强制逻辑校验（如日期格式、数值范围）；
- e) 院外采集的生物样本（如留存血清）需同步生成匿名化条码。

4.5.2 技术保障

4.5.2.1 数据采集工具标准

数据采集工具的要求包括：

- a) 不同研究地区宜保持一致的信息收集维度、指标定义标准；
- b) 宜采用计算机辅助面访或计算机辅助自填问卷的形式，以保证研究数据的质量。

4.5.2.2 电子数据收集系统

电子数据收集系统的要求包括：

- a) 实现与医院信息系统的数据接口，完成数据自动抓取、预处理、标准化格式转换；
- b) 应具备逻辑检错、时间戳记录等功能，确保可追溯性与一致性。

4.5.3 人员管理

4.5.3.1 登记员要求

登记员是队列招募流程中与研究对象接触的第一个岗位，意义重大，调查员应满足以下要求：

- a) 经过统一培训，熟悉登记的流程与调查对象入选规则；
- b) 熟悉当地方言，态度和蔼，善于沟通，有亲和力，具备现场问题处理能力；
- c) 举止稳妥，工作耐心。

4.5.3.2 数据收集员要求

数据收集员的要求包括：

- a) 经过统一培训，考核合格后方能上岗；
- b) 熟悉信息收集各维度的含义；
- c) 使用计算机辅助问卷的调查员应有一定的计算机基础，能熟练操作调查所需的设备；
- d) 熟悉当地语言，能用当地语言与调查对象正常交流，熟悉当地生活方式和习惯；
- e) 善于沟通。耐心和蔼，头脑灵活，能控制交谈局面。

4.6 质量控制

为确保队列收集数据的真实性、完整性与可追溯性,需构建覆盖数据采集全流程的质量控制与评估体系。质量控制内容包括:过程管理、数据采集质控、质控指标。

4.6.1 过程管理

4.6.1.1 培训

针对新上岗参与人员招募和数据收集工作的相关人员应进行技术培训,对于承担调查工作的业务人员也应定期开展集中培训。培训内容包括调查时间、内容、方法(流程)、质控要求及评价指标等。

4.6.1.2 例会

工作中应定期召开工作例会,核心内容包括核对调查名单及相关记录,讨论发生终点事件及调查有困难的研究对象等;交流长期调查过程中的工作经验以及反馈调查过程中存在的问题以及改进措施。

4.6.1.3 督导

应抽取一定比例研究对象的终点事件报告卡进行填写质量检查,并检查名单等各类资料的整理、填写与记录情况。

4.6.1.4 考核

应定期对承担调查工作的单位和个人进行考核,应设定质控指标进行量化考核,并对存在的问题及时协调与处理。质控指标选择以考核终点事件的及时、完整与准确为主(见4.6.3评价指标)。

4.6.2 数据采集质控

4.6.2.1 采集工具数据校验

在电子问卷系统中嵌入逻辑校验规则(如逻辑跳转、数值范围限制),对填写错误提供实时提示或禁止录入,避免低级错误。针对关键字段设置必填项及时间戳机制,确保记录完整可追溯。

4.6.2.2 录音审核

在调查的同时录音,并对照录音评价询问是否符合要求。

4.6.2.3 随机复核

随机抽取一定比例的研究对象,进行数据复查与信息比对,验证一致性与可靠性。

4.6.2.4 现场过程督导

安排监督员在调查现场实时观察或事后回顾资料,确保调查过程规范执行,避免数据造假、漏报等情况。

4.6.3 评价指标

4.6.3.1 及时性

评价终点事件收集的及时性，主要指从研究对象发生死亡、感染性休克等疾病、出院或再入院等终点事件的时间到录入队列数据库的时间。

4.6.3.2 完整性

评价终点事件数量收集的完整性（如核对随访数据库中是否有漏报）以及终点事件报告卡填报的完整性。

4.6.3.3 准确性

评价终点事件报告卡填写的准确性（如信息填写是否有逻辑错误，死亡信息是否有死因不明、死因不正确或不符合推导规则等）以及终点事件收集的准确性。

4.6.3.4 一致性

评价终点事件报告的一致性和诊断的一致性，以及常规监测、社区定向监测之间的对比和复核。

4.6.3.5 失访率

在队列人群长期随访中，无法收集到终点事件信息的研究对象人数占队列人群总数的比例。

4.7 伦理审查和人类遗传资源管理

4.7.1 伦理审查

为避免对人类尊严、人权和基本自由等造成的伦理风险，在感染性休克队列构建及临床研究等项目工作开始前必须进行完整的伦理审查。伦理审查需要向具有资质的伦理委员会提交申请和报告，并按照委员会相关要求提交审查材料。伦理审查通过前，不得开始任何现场工作。在队列构建及研究运行过程中，应遵守伦理委员会的规定，并按期向伦理委员会提交持续审查申请。到项目结束时，须提交项目结题申请。如出现意外情况，须及时报告伦理委员会，并提交严重不良事件报告，视情况提交暂停或终止研究申请。对于实施过程中出现的研究方案变动，须提交修正案申请并通过后，方能继续进行现场工作和数据收集。

4.7.2 人类遗传资源管理

对于感染性休克队列中涉及的生物样本和相关数据，必须遵守《人类遗传资源采集、收集、买卖、出口、出境审批》、《人类遗传资源管理暂行办法》、《人类遗传资源管理条例》等相关法规条例进行管理，规范和促进相关人类遗传资源的保护和利用。同时，对于队列中重要的遗传家系和特定地区遗传

资源，要及时向有关部门申报和登记。

4.7.2.1 生物样本接收

生物样本接收的具体要求如下：

- a) 人类生物样本库建库应获得中华人民共和国人类遗传管理办公室行政许可批件；
- b) 样本接收应根据样品的特性制定样品接收程序文件，并制定接收/拒收样品的标准，确保样本质量在接收活动中不受损害；
- c) 应确保接触样本的人员接受安全操作规程培训；
- d) 工作人员应保护所接触供体、样本及相关数据的隐私；
- e) 应由专业人员在符合要求的指定场所进行；
- f) 应具备采集单位提供的知情同意及伦理审批文件。

4.7.2.2 生物样本储存

生物样本储存的具体要求如下：

- a) 应确保样本储存的设备及耗材具有唯一标识；
- b) 应配备适量的备用储存设备；
- c) 应建立储存温度失控后的应急处理预案；
- d) 样本储存应建立入库与出库程序文件，按程序文件实施入库并记录过程信息；
- e) 应根据样本类型和预期用途选择合适的储存条件和设备；
- f) 宜利用信息化系统，对样本的出入库和样本的储存环境（温度、湿度）及位置等进行监控管理；
- g) 应制定库存核实计划，盘点整理库存样本并记录。

4.7.2.3 生物样本分发

生物样本分发的具体要求如下：

- a) 样本及相关数据应匿名化；
- b) 样本分发应提前告知申请人样本分发的时间、地点及需要准备的包装和运输材料；
- c) 应在合适的条件下进行分发活动，确保样本的质量；
- d) 样本及相关数据分发前，应与申请人签订样本和/或数据转移协议。

4.7.2.4 生物样本包装

生物样本包装的具体要求如下：

- a) 样本包装应根据样本的运输温度要求，采用已验证的温控包装。对运输过程中的温度状况，进行实时监测和控制，保证样本的运输温度控制在规定范围内；

- b) 应将包装样本容器内的空余位置进行填充;
- c) 对温度有要求的样本应准备合适和足够的制冷剂;
- d) 对湿度有要求的样本应将样本置于有干燥剂的密封袋中;
- e) 对光敏有要求的样本应将样本置于遮光的密封袋中。

4.7.2.5 生物样本运输

生物样本运输的具体要求如下:

- a) 样本运输应符合相关安全规定和相关标准的要求。如实申报运输样本以及制冷剂的内容和潜在的危险性;
- b) 应由合格的工作人员或委托有资质的物流公司承运, 根据承运方的资质和条件宜对承运方的相关人员进行培训;
- c) 根据样本类型、运输距离、时间及样本的温度、外部环境温度等情况, 选择合适的运输方式和温控包装, 确保样本在运输过程中符合温度控制要求;
- d) 样本运输前应准备所有运输文件。运输文件应包括样本相关信息及运输相关的文件和证明等;
- e) 样本运输开始前样本的发送方和接收方应进行沟通, 保证接收方知晓样本可能到达的时间并做好接收准备;
- f) 样本的发送方和接收方都应对样本的运输进行追踪, 以便出现问题时能及时解决;
- g) 全程温度数据、操作节点时间、操作人员等监管数据不可删除和更改, 且可导出或上传。

4.7.2.6 生物样本弃用

生物样本弃用的具体要求如下:

- a) 样本弃用(包含样本退回和销毁)应遵循伦理相关要求;
- b) 应遵循国家法律、法规要求, 建立样本弃用管理程序文件并记录实施过程信息;
- c) 应确保样本及相关数据在弃用或销毁的全过程具有可追溯性。

符合以下条件之一可申请样本弃用:

- a) 符合研究方案中弃用要求;
- b) 样本的采集量超出已批准方案限定的范围;
- c) 超出知情同意书限定的范围;
- d) 供体要求撤销样本或退出研究等;
- e) 样本相关数据丢失;
- f) 样本受到污染或破损;
- g) 有新的研究表明样本具备潜在的风险和生物危害;
- h) 保藏过程导致的样本不可用。

5 数据要素构成

5.1 电子病历数据

感染性休克电子病历标准数据集各模块主要参考国家电子病历及信息化行业标准以及最新脓毒症和感染性休克领域诊疗指南。数据集共 13 个模块分类，具体的数据元字段如下：

（1）基本信息

包括患者标识、性别、出生日期、年龄、民族、身高、体重等。

（2）就诊记录

包括入院日期、出院日期、入院科室、出院科室、就诊日期、就诊科室等。

（3）现病史

包括患者主诉信息等。

（4）既往史

包括既往相关疾病、是否药物治疗、药物名称、用药剂量、用药时间等。

（5）个人史

是否有吸烟史、吸烟年数、平均每日吸烟数量、最近一次戒烟日期；是否有饮酒史、饮酒类型、饮酒年数、平均每日饮酒总量、最近一次戒酒日期等。

（6）家族史

包括家族疾病名称、家族疾病亲属关系等。

（7）体格检查

包括检查部位、检查结果、异常描述等。

（8）诊断

包括诊断 ICD-10 编码等。

（9）实验室检查

包括检验日期、检验项目名称、检验定性结果、检验定量结果、检验定量结果单位、检验结论。

检验细类包括血常规（红细胞计数、血红蛋白、红细胞压积、平均红细胞体积、平均红细胞血红蛋白（HGB）含量、平均红细胞 HGB 浓度、红细胞分布宽度标准差、红细胞分布宽度变异系数、血小板计数、白细胞计数、中性分叶核粒细胞百分率、淋巴细胞百分率、单核细胞百分率、嗜酸性粒细胞百分率、嗜碱性粒细胞百分率、原始细胞百分率、中性分叶核粒细胞绝对值、淋巴细胞绝对值、单核细胞绝对值、嗜酸细胞绝对值、嗜碱细胞绝对值）；血生化（总胆红素、直接胆红素、间接胆红素、丙氨酸氨基转移酶（ALT）、门冬氨酸氨基转移酶（AST）、总蛋白、白蛋白、球蛋白、葡萄糖、尿素、肌酐、估算肾小球滤过率、血清胱抑素 C 测定、尿酸、甘油三酯、胆固醇、高密度脂蛋白、低密度脂蛋白、碱性磷酸酶、谷氨酰转肽酶、肌酸激酶、乳酸脱氢酶、羟丁酸脱氢酶、钠、钾、氯、二氧化碳结合力、阴

离子间隙、血清 β 羟基丁酸测定、血钙、血镁、血清无机磷、总胆汁酸）；血气分析（动脉 PH 值、氧分压、二氧化碳分压）等。

（10）床旁实时指标收集

包括体温（ $^{\circ}\text{C}$ ）、收缩压（mmHg）、舒张压（mmHg）、心率（次/分）、呼吸频率（次/分）、吸氧浓度（%）、氧饱和度（%）、睁眼反应、是否插管、言语反应、运动反应等。

（11）综合性评价指标

SOFA、qSOFA、APACHE II 评分等。

（12）药物治疗

包括治疗开始时间、治疗结束时间、药物名称、药物通用名药物单次剂量、药物剂量单位、给药途径、用药频次等。

（13）疗效评价

包括评估日期、评估方法、总体疗效评估等。

5.2 医学影像

5.2.1 超声影像

5.2.1.1 数据格式与存储

数据文件格式应包括：DICOM、JPEG、BMP 等。如有射频（RF）信号数据，亦可保存为二进制 RF 格式。

5.2.1.2 数据质量标准

最低分辨率要求应为 72ppi，图像尺寸应为 1024 像素×768 像素，图像文件大小应不小于 768kb。图像清晰、均匀性、超声切面标准、图像无斑点，雪花细粒，网纹等干扰。

5.2.2 数字组织病理影像

5.2.2.1 数据格式与存储

数据文件格式应包括：Generic tiled TIFF、JPEG、BMP、JPEG2000 等。

5.2.2.2 数据质量标准

质量要求包括：

- a) 切片完整情况：切片完整、无破损情况；
- b) 切片染色情况：染色保持鲜亮，视觉感知明显，能清晰辨识细胞大小及形状；
- c) 细胞以及核边界清晰；
- d) 包含细胞分布区域 95% 以上；

e) 未引入涂片不包含的图像噪声，无外部噪声影响阅片诊断。

5.2.3 MRI

5.2.3.1 数据格式与存储

数据文件格式应包括：DICOM、JPEG、BMP 等。

5.2.3.2 数据质量标准

质量要求包括：

- a) 应使用原厂水模和规程定期对磁共振成像设备性能进行质控；
- b) 常规图像无明显畸变，弥散加权图像畸变在正常范围内；
- c) 图像信号均匀，无明显信噪比异常，无明显射频干扰及其他各种形式的伪影；
- d) 患者配合良好，图像无明显运动伪影；
- e) 图像平面内分辨率应满足临床诊断要求；3D 头部扫描层厚不超过 1 mm，头部普通扫描层厚不超过 3 mm；心脏、腹部扫描层厚不超过 5 mm；关节、软骨扫描层厚不超过 3 mm。磁共振检查设备建议说明 1.5 T 及以上；扫描层厚要求过薄，泛化推广欠佳，大量临床数据将被排除在外；建议增加层间距的参数；建议增加图像质量评分标准。

5.2.4 CT 影像

5.2.4.1 数据格式与存储

数据文件格式应包括：DICOM、JPEG、BMP 等。

5.2.4.2 数据质量标准

（1）一般图像质量要求

每个病例的图像应当保持连续完整，不得出现缺层、错层等情况。不得进行任何修改、编辑，不得进行有损压缩。根据对解剖细节的显示情况，分为以下几类：

- a) 可见：器官和结构在检查范围内可观察到。但细节未显示；
- b) 显示：解剖结构细节可见，但不能清晰辨认，即细节显示；
- c) 清晰显示：解剖细节清晰辨认，即细节清晰。

（2）冠脉图像质量要求

可诊断的冠状动脉（直径 ≥ 1.5 mm）节段中的 90%（15 个节段中的 13 个节段）没有伪影，能够诊断；确保冠状动脉和心脏扫描范围的完整性；冠状动脉 CT 值最佳范围 300 HU~450 HU，特别注意冠状动脉远端增强效果是否满意；图像噪声：测量主动脉根部图像的 CT 值标准差（SD 值）作为图像噪声， <20 HU 为优秀， 20 HU~ 30 HU 为良好， >30 HU 为图像质量差， >40 HU 为检查失败（图像不

能评估）。推荐目标控制在 20 HU~30 HU 以下，推荐使用迭代重建降低图像噪声。

5.2.5 X-ray 影像

5.2.5.1 数据格式与存储

数据文件格式应包括：DICOM、JPEG、BMP 等。

5.2.5.2 数据质量标准

质量要求包括：

- a) 清晰可见：解剖学结构或/和病变特征细节可清晰辨认，即细节清晰；
- b) 可见：解剖学结构或/和病变特征等细节可见，但不能清晰辨认，即细节显示；
- c) 隐约可见：解剖学结构或/和病变特征等细节可观察到，但细节没有完全显示，只特征可见。

5.2.6 心电图

5.2.6.1 数据格式与存储

数据文件格式应包括：BMP、JPEG、PDF、FDA-XML（HL7）等。

5.2.6.2 数据质量标准

能分辨小波形与伪差之间的鉴别等细节；能识别 P、QRS、T、U 波形及各个间期，如 PR 间期、ST 段、QT 间期；扫描电子病历和纸质病历心电图图像应清晰体现心电图标尺，即每小格表示的振幅（纵坐标）及时长（横坐标）。

5.2.7 脑电图

5.2.7.1 数据格式与存储

数据格式与存储应符合下列要求：

- a) 脑电图数据文件格式应包括：EDF、TXT、MAT、XLS 等，数据内容应包括各通道电极所记录的数据以及电极类型、数量、位置、固定方式、阻抗等附加信息；
- b) 采用脑电和近红外光同时记录时，近红外光部分数据格式为三维矩阵 $Pa \times Ch \times Nt$ ，其中 Pa 代表 fNIRS 检测的参数种类，至少包含含氧血红蛋白与脱氧血红蛋白两种参数， Ch 代表通道数， Nt 代表时间，建议大于 100。文件类型应包括 TDMS，XLS，MAT 和 TXT。

5.2.7.2 数据质量标准

- a) 系统增益与分辨率：信号放大器应具有至少 100 dB 的共模抑制比以消除共模噪声。分辨率不应低于 8 bit。

- b) 滤波要求: 应记录带通滤波器的上、下限截止频率 (-3 dB)。转换器采样频率应高于事件相关电位 (ERP) 信号最高频率的两倍。通常模拟滤波器上限截止频率可设定为转换器采样频率 1/4, 下限截止频率为扫描时间 4 倍的对应频率。如果需要进一步滤波, 可用数字滤波进行后续分析。滤波器并不能完全滤除截止频率之外的频率成分, 超出滤波范围的强信号仍可能导致混叠现象。此外, 截止频率 50 Hz~60 Hz 的陷波滤波器可能会使脑电事件相关电位波形严重失真, 应谨慎使用;
- c) 采样率要求: 模数转换器的采样频率会影响测量信号的拾取效果, 应记录标明具体采样频率。采样频率不应低于所测量信号最高频率成分的 2 倍, 以防止发生混叠现象, 造成波形失真;
- d) 信号伪迹和污染数据剔除要求: 从头皮记录的电信号并不是单纯的脑电, 头部肌肉运动、舌动、眼动、皮电等噪声都对该信号有影响。信号伪迹可包括: 生理性伪迹 (如: 心电伪迹、肌电伪迹、眼动伪迹等)、来自采集系统的伪迹、电磁干扰伪迹、运动伪迹等。为便于脑电后续分析去除伪迹, 降低其他源信号对脑电信号的干扰, 应设置其他信号 (如: 心电、肌电等) 的检测电极并准确记录并报告采用的电极数量、位置、作用等信息。可根据脑电信号参数 (例如: 潜伏期、幅值等) 判定数据受污染程度, 一般情况下如果某单次数据明显超出 ($\pm 200 \mu\text{V}$) 该通道的叠加平均, 则该次数据不应纳入叠加平均。根据测量信号种类不同, 该剔除标准也会发生改变。应记录不同实验条件、不同受试所剔除序列的比例。对于容易产生不合格数据的受试群体, 宜增加数据记录数量;
- e) 光电联合成像时的数据评估要求: 可通过主观观察或者算法提取原始静息态数据中的脉搏信号: 如果能够清晰观察或提取到与脉搏频率一致的波动信号, 则可认为数据优良; 如果观察到或提取到信号波动不明显, 则可认为数据中等, 如果无法提取或观察到脉搏信号即为数据较差。应取优良信号和中等信号做分析, 其相应占比宜为 90% 和 10%。

5.3 组学数据

可按照研究目标进行设计, 纳入不同类型组学数据, 包括但不限于基因组、转录组、蛋白组、代谢组、单细胞组等。数据格式、存储及质控等内容可参考《GB/T 30989-2014 高通量基因测序技术规程》、《GB/T 35537-2017 高通量基因测序结果评价要求》、《GB/T 35890-2018 高通量测序数据系列格式规范》、《GB/T 33767.14-2023 信息技术 生物特征样本质量 第 14 部分: DNA 数据》、《GB/T 45214-2025 人全基因组高通量测序数据质量评价方法》、《T/CMEAS 045-2025 脓毒症多组学数据库建设及应用》等标准相关内容执行。

5.4 数据标准化存储

感染性休克队列电子病历及影像数据存储标准可参考下表:

感染性休克队列标准数据集

数据名称	数据大小（字节）
基本信息	78
就诊记录	75
现病史	72
既往史	610
个人史	56
家族史	90
体格检查	1247
诊断	21
实验室检查	306
床旁实时生命体征	221
药物治疗	111
疗效评价	216

数字组织病理影像数据	
数据名称	数据大小（字节）
超声影像	768K
数字组织病理影像	1G
MRI	25M
CT 影像	150M
X-ray 影像	8M
心电图	13.5M
脑电图	230M

6 数据汇集

6.1 数据采集设备

6.1.1 兼容性要求

6.1.1.1 电子病历数据

兼容医院内各类数据接入源，包括原业务数据库的备份库、数据中心、集成平台和文件接口。数据库方面支持 Oracle、MySQL、SQL Server 等主流数据库平台。

6.1.1.2 医学影像数据

可参照《T/CMDA 001-2020 肝胆疾病标准数据规范：肝癌 CT/MRI 影像采集和处理标准》等标准文件，如使用不低于 16 排的螺旋 CT；使用不低于 1.5 T 的 MRI（体部相控阵线圈），有医疗器械注册证的国内外厂家等。

6.1.2 采集要求

6.1.2.1 采集方式

在院内部署数据采集工具，对接院方信息科提供的原业务数据库的备份库、数据中心、集成平台或者文件接口。

6.1.2.2 采集协议

采用 HTTP 协议或者数据库连接协议来采集标准的电子病历和医学影像数据。

6.1.2.3 采集参数

按照感染性休克病例数据标准要求的数据参数指标进行采集。

6.1.2.4 采集精度

按照感染性休克病例数据标准要求的数据精度进行采集。

6.2 数据采集过程

6.2.1 采集人员要求

6.2.1.1 选拔

经过医院信息科和相关部门同意的人选。

6.2.1.2 培训

对上岗进行数据采集前的操作人员进行培训，培训内容包括：采集设备使用、采集方式、采集参数及精度、采集质量要求等。

6.2.1.3 考核

在采集过程和结果中，按照相关标准进行考核。

6.2.2 采集过程要求

6.2.2.1 采集准备

采集之前，在医院部署相应的物理服务器，在服务器上部署相应的采集工具，与医院信息科商定采集接口，这部分工作完成即已做好采集前的准备工作。

6.2.2.2 采集过程

采集过程严格按照相关安全标准进行。

7 数据治理

7.1 数据脱敏

7.1.1 类型

数据集使用的数据必须是获得医院伦理委员会批准或者豁免的临床脱敏数据，患者的隐私保护应当满足法规的要求，保证数据脱敏、患者隐私安全和患者利益，故数据应严格遵循法规要求进行脱敏处理。

7.1.2 内容

被脱敏的信息包括：

- a) 电子病历中病人姓名、年龄、性别、ID号、检查日期、出入院日期等所有涉及敏感信息的数据；
- b) 医学影像数据中涉及敏感信息的标签等数据内容。

7.1.3 方法

确认具体敏感字段之后，针对相关的信息以下面的方式进行脱敏操作。

7.1.3.1 姓名的去标识化

姓名是一种常用的标识符，适用的去标识化方法举例如下：

- a) 泛化编码：使用概括、抽象的符号来表示，如使用“张先生”来代替“张三”，或使用“张某某”来代替“张三”。这种方法是用在需要保留“姓”这一基本特征的应用场景。
- b) 抑制屏蔽：直接删除姓名或使用统一的“*”来表示。如所有的姓名都使用“****”代替。
- c) 随机替代：使用随机生成的汉字来表示，如使用随机生成的“辰筹猎”来取代“张三丰”。
- d) 假名化：构建常用人名字典表，并从中选择一个来表示，如先构建常用的人名字典表，包括龚小虹、黄益洪、龙家锐等，假名化时根据按照顺序或随机选择一个人名代替原名。如使用“龚小虹”取代“张三丰”。这种方法有可能用在需要保持姓名数据可逆变换的场景。

e) 可逆编码: 采用密码或其他变换技术, 将姓名转变成另外的字符, 并保持可逆特性。如使用密码和字符编码技术, 使用“SGIHLIKHJ”代替“张三丰”, 或使用“FzF”代替“Bob”。

7.1.3.2 地址的去标识化

对于地址, 常用的去标识化方法举例如下:

- a) 泛化编码: 使用概括、抽象的符号来表示, 如“江西省吉安市安福县”使用“南方某地”或“J省”来代替。
- b) 抑制屏蔽: 直接删除地址或使用统一的“*”来表示。如所有的地址都使用“*****”代替。
- c) 部分屏蔽: 屏蔽地址中的一部分, 以保护地址信息。如使用“江西省 xx 市 xx 县”来代替“江西省吉安市安福县”。
- d) 数据合成: 采用重新产生的数据替代原地址数据, 数据产生方法可以采用确定性方法或随机性方法。如使用“黑龙江省鸡西市特铁县北京路 23 号”代替“江西省吉安市安福县安平路 1 号”。

7.1.3.3 数值型标识符的去标识化

常用的数值型标识符的去标识化包括:

- a) 泛化编码: 使用概括、抽象的符号来表示, 如“有四个人, 他们分别是蓝色、绿色和浅褐色的眼睛”来代替“有 1 个人是蓝色眼睛, 2 个人是绿色的眼睛, 1 个人是浅褐色的眼睛”。
- b) 抑制屏蔽: 直接删除数值或使用统一的“*”来表示。如所有的数值都使用“*****”代替。
- c) 顶层和底层编码: 大于或小于一个特定值的处理成某个固定值。例如, 年龄超过 70 岁的一律用“大于 70 岁”描述, 以保障满足此条件的人数多于 20000 人。
- d) 部分屏蔽: 使用数值的高位部分代替原有数值, 如百分制考试成绩全部使用去掉个位数、保留十位数的数值代替。
- e) 记录交换: 使用数据集中其他记录的相应数值代替本记录的数值。如设定规则, 将记录集中的所有的身高数据取出并全部打乱位置后 (其他属性数据位置不变) 放回原数据集中。这种方法可以保持数据集的统计特性不变。
- f) 噪声添加: 相对原始数据, 产生微小的随机数, 将其加到原始数值上并代替原始数值。如对于身高 1.72m, 产生随机数值-0.11m, 加到原始数值后将其变为 1.61m。
- g) 数据合成: 采用重新产生的数据替代原始数据, 数据产生方法可以采用确定性方法或随机性方法。如使用“19”岁年龄代替“45”岁年龄。

7.1.3.4 日期去标识化

在数据集中, 日期有多种存在形式, 包括出生日期、开始日期、纪念日等。常见的对日期的去标识化, 方法包括:

- a) 泛化编码：使用概括、抽象的日期来表示，如使用 1880 年代替 1880 年 1 月 1 日。
- b) 抑制屏蔽：直接删除日期数据或使用统一的“*”来表示。如所有的数值都使用“某年某日”代替。
- c) 记录交换：使用数据集中其他记录的相应数值代替本记录的数值。如设定规则，将记录集中所有的日期数据取出并全部打乱位置后（其他属性数据位置不变）放回到原数据集中。这种方法有利于保持数据集的统计特性。
- d) 噪声添加：相对原始数据，产生微小的随机数，将其加到原始数值上并代替原始数值。如对于出生日期 1880 年 1 月 1 日，产生随机数 32 天，加到原始数值后将其变为 1880 年 2 月 2 日。
- e) 数据合成：采用重新产生的数据替代原日期数据，如使用“1972 年 8 月 12 日”代替“1880 年 1 月 1 日”。

7.1.3.5 医学影像去标识化

常见的医学影像去标识化方法包括：

- a) 图像遮掩：通过对医学影像进行语义分割，精准识别并提取包含敏感信息的区域，继而应用掩码对其进行覆盖或模糊处理，从而实现图像层面的有效脱敏。
- b) 替换：识别医学影像中包含的特定敏感字段（如患者姓名、ID 号、医院名称等），并对这些字段进行屏蔽或替换处理，例如使用“*”替代原始信息，以达到文本内容的匿名化目的。

7.1.4 流程

脱敏管理流程见图 1。

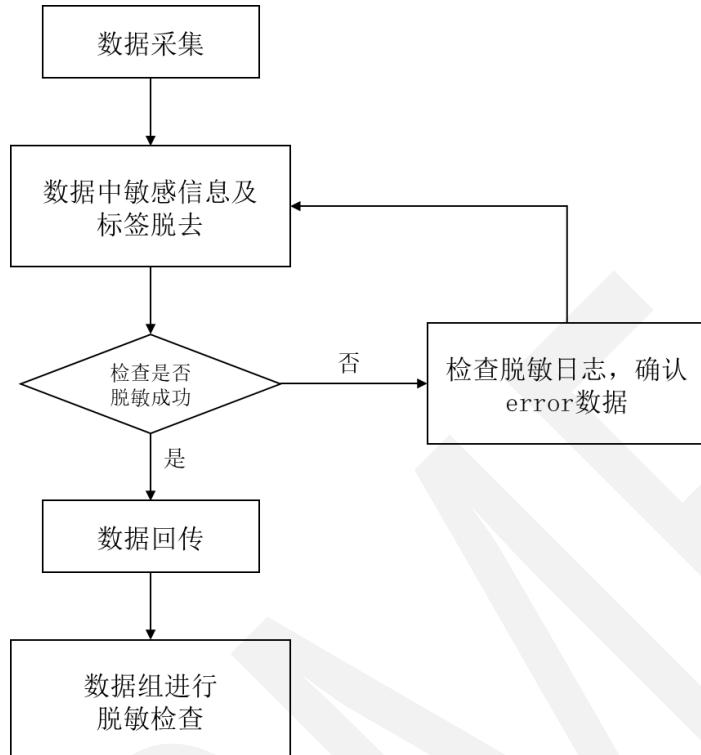


图 1 脱敏管理流程

7.2 数据处理

7.2.1 规则

7.2.1.1 清洗规则

(1) 电子病历数据

电子病历数据清洗规则包括非空检核、主键重复、非法代码清洗、非法值清洗、数据格式检核、记录数检核及业务约束核验等。

- a) 非空检核：要求字段为非空的情况下，需要对该字段数据进行检核。
- b) 主键重复：多个业务系统中同类数据经过清洗后，在统一保存时，为保证主键唯一性，需进行检核工作。
- c) 非法代码、非法值清洗：非法代码问题包括非法代码、代码与数据标准不一致等，非法值问题包括取值错误、格式错误、多余字符、乱码等，需根据具体情况进校核及修正。
- d) 数据格式检核：通过检查表中属性值的格式是否正确来衡量其准确性，如时间格式、币种格式、多余字符、乱码。
- e) 记录数检核：指各个系统相关数据之间的数据总数检核或者数据表中每日数据量的波动检核。
- f) 业务约束核验：在实施处理数据过程中，与医院相关人员共同确定。从业务的正确性、一致性、

有效性等角度，考虑数据的核验规则，如：建档日期、入院日期、民族信息等的有效性核验。

（2）医学影像数据

医学影像数据清洗规则主要包括检查数据扫描质量，数据是否存在跳层、少层问题，存在伪影遮挡问题；检查数据扫描完整性，所要求的序列有无缺失、有无超范围扫描。

7.2.1.2 脏数据处理

（1）电子病历数据

对于常见的空缺值、离群值和不一致等脏数据，采用人工检测、统计学方法、聚类、分类、基于距离、关联规则等方法来实现数据清洗。

根据缺陷类型分类，可以将脏数据分为缺失值数据、错误数据和错误关联数据这三种典型问题数据，分别进行数据清洗。

（2）医学影像数据

医学影像数据处理主要包括以下步骤：

- a) 数据去重：删除重复的图像；
- b) 数据过滤：删除低质量、不标准、不合理的图像；
- c) 数据修复：修复数据中出现的缺失、噪声、伪影等问题；
- d) 数据增强：增加数据的数量和多样性，例如对图像进行旋转、平移、缩放、翻转等操作，以及在图像中添加噪声等。

7.2.2 方法

（1）电子病历数据

通过人工复核按照下面的方式进行验证：

- a) 初期的数据要求提供数据报告，包括接入数据范围（系统数量）、数据量（需要分类，比如患者、医嘱、处方、检验检查数量等）、数据纵深（起始到终止日期）。医院信息部门可进行宏观数据核验。
- b) 数据质量抽样检查：比照医院信息系统（HIS）、电子病历系统（EMR）等业务系统的单个患者进行数据核验。
- c) 病案数据抽样检查：随机抽取一定数量的病例数据进行核对。
- d) 隐私数据检查：查看是否按照要求进行去隐私化处理，比如患者姓名、地址、身份证号等，可有意识的搜索隐私数据看可否能检索到。检出问题后及时汇总和分析出问题原因。要么接入对接偏差，要么工程师理解偏差，或是后期后结构化归一的次生原因，作为后期调整技术方案的依据。

（2）医学影像数据

数据扫描质量和完整性检查程序通过读取元数据与真实采集数据进行比对，通过检查规则设定判

断数据跳层、少层、序列缺失、超范围扫描、伪影遮挡问题。伪影遮挡问题除了程序自动化判断外，需要人工复核。

7.2.3 通过准则

以下 8 个质量维度进行整体评估：

- a) 完整性：数据项内容的完整情况，主要考察字段数据是否存在，NULL 等同于空；
- b) 规范性：数据项的数据格式，数据长度，时间范围，数值范围是否正确；
- c) 标准性：数据项的数据是否与字典一致，如性别的取值应该是男，女；
- d) 正确性：数据项内容是否正确，如入院记录中月经史患者性别不应该为男；
- e) 一致性：数据项间内容是否一致，如考察病案首页上的患者、手术、诊断、费用等信息是否同 HIS 或其他系统中的信息一致；
- f) 时序性：时间逻辑合理性（流程时间分布），如出院时间需要晚于入院时间；
- g) 整合性：相关系统对应数据项目可对照或关联，如医疗费用明细表里的医嘱标识应该在中西医处方表、药品类医嘱表、草药处方表或者非药品医嘱表里；
- h) 唯一性：记录的唯一性，如 Patient ID 在 patient 表不能重复，相同的医嘱不应该重复。

7.3 数据管理

7.3.1 人员管理

7.3.1.1 选拔

医院信息和相关部门同意的人选。

7.3.1.2 培训

信息部门或者大数据其他主管部门的培训，主要包含用户授权、软件安装、数据加密和反加密工具使用、数据质量监控、数据接入监控、数据统计工具使用、数据导出管理、标注内容培训等。建议入选的标注医师接受相关培训，同意对标注规则的认识，熟悉标注软件操作。

7.3.1.3 考核

通过制定相关的考核制度进行考核，内容包括数据安全管理制度、数据使用管理规范、使用违规惩罚制度、标注规范等四部分：

- a) 数据安全管理制度：数据安全管理从系统安全和审批管理两方面进行考虑。系统安全包括信息部门数据备份策略、用户权限、账户弱密码、账户有效期和数据导出管理等方面进行管理；审批管理主要指账户权限审批流程和数据申请审批流程；
- b) 数据使用管理规范：数据使用人需严格遵守国家有关法律法规，采集数据时需注明申请数据范

围、用途、使用事件等重要信息，对使用数据有保密责任，数据采集结束要立即向管理部门报备，不可随意使用数据或进行其他违规操作；

- c) 使用违规惩罚制度：对于擅自传播、转让、更改数据用途、未经许可使用数据或其他违规行为，医院管理部门有权利追究法律责任；
- d) 标注规范。

7.3.2 基础设施管理

7.3.2.1 场所

在医院内部网络对数据进行管理，未经管理部门批准不连接互联网。

7.3.2.2 软件

使用稳定的专业医学软件进行数据管理工作。

7.4 数据标注

7.4.1 标注人员

为提高感染性休克临床指标标注的客观性和一致性，标注工作需要 2 名主治以上级别的医师参加。每一例标注任务由两名医师达成一致意见才能确定，否则由一名仲裁专家介入、进行评定。仲裁专家由工作经验 15 年以上的副主任医师或主任医师担任。

7.4.2 标注流程

在数据标注过程中需要确认标注的目标以及具体的标注条件。以下以病历报告中的脓毒症标注作为样例（图 2），具体数据标注需求和条件请专家结合领域知识设计。考虑到本标准涉及较多的数据模态，在必要情况下需要针对模态设计对应的标注要求和标注流程规范。

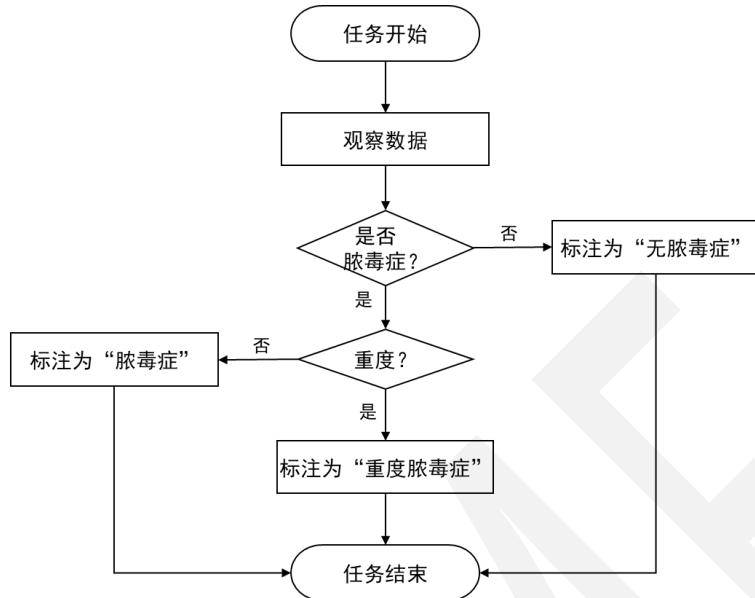


图 2 脓毒症标注流程示例

7.4.3 标注任务

不同数据类型及需求，具体标注任务包括如下方面。

7.4.3.1 分类标注

即将数据按照人工判读得到的数据类别进行标签标记，通常可分为二分类标记和多分类标记，例如通过解读病历文本进行“无脓毒症”、“有脓毒症”（二分类）或“脓毒症”、“严重脓毒症”、“脓毒性休克”（多分类）标记；

适用：病历文本、医疗影像数据。

7.4.3.2 区域标注

对数据拟标注目标的范围进行标注，如对病历文本中的实体起始字符位置进行标记或对医疗图像中目标区域范围进行标记；

适用：病历文本、医疗影像数据。

7.4.3.3 标框标注

框选标注需要检测的目标对象，如在 CT 图像中框选出感染位置；

适用：医疗影像数据。

7.4.3.4 描点标注

对于细致特征的要求中需要将拟检测目标进行描点标记，如描点标记出 CT 图像中的感染部位具体形状；

适用：医疗影像数据。

7.4.3.5 光学字符识别（OCR）转写标注

识别图片格式中的手写文字，转写为计算机可识别的文本格式，如识别图片扫描格式的病历文本；

适用：图片扫描的病历文本。

7.4.4 标注细则

标注过程中应尽量做到无错标、漏标，对无法确定具体类别的数据样本纳入集中管理并呈交专家组进行标记和复审。对于图像标记应做到以不同颜色区分主要征象和次要征象，以标记主要征象为主、尽量多标次要征象；在标注病灶轮廓时，对内部细节辅以文字进行描述。

7.4.5 分歧处理

在标注过程中应尽可能保证客观性，如在此前提下仍无可避免地存在一定主观性（比如对某一疗效好坏的判断需要由标注医生根据病历结合经验进行主观判断），致使标注医师的标注结果存在标准容许以外的差异性，则可由专家组进行标注结果的仲裁和判定。

7.4.6 自动标注

允许利用标注系统自动进行标注，但其结果仍需医疗专家审核和评价。

7.4.7 标注质控

7.4.7.1 人员

仲裁专家由从事重症医学领域工作经验 15 年以上的副主任医师或主任医师担任。

7.4.7.2 方法

为提高标注的准确性，建议将标注流程进行多轮次、分组、交叉设计，整个标注工作由多级别医师分组（标注医师、标注组长、仲裁专家）参与。标注质控流程（图 3）建议如下：

- (1) 数据标注工作由标注组长带领至少 3 名标注医师独立完成，可在标注数据中设置一定数量的重叠样本以检测标注一致性；
- (2) 标注完成后，先由计算机对标注结果进行自动比较和归集，对于标注一致的结果进行合并，对标注不一致的结果以特殊颜色提醒仲裁专家进行标注；
- (3) 由标注组长对标注结果进行复审、仲裁专家进行终审从而形成最终标注结果。

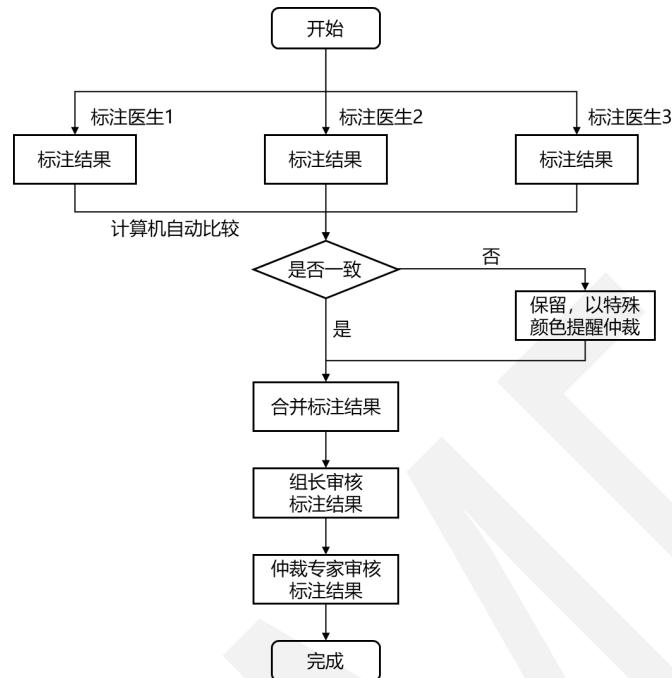


图3 标注质控流程图示

7.4.7.3 指标

针对感染性休克疗效评价，普通病历的评价由3名标注医师执行，由1名审核组长进行审核，1名仲裁专家进行仲裁。遇到复杂征象时，可酌情增加审核人数，以保证标注质量，审核组长和仲裁专家可以手动修改。

7.4.7.4 通过准则

标注评估的通过准则为：

- 检出环节：3名标注医师背靠背独立标注，然后用计算机自动判断检出的一致性，以所有人标注结果的并集作为结果。
- 审核环节：由其他标注组长和仲裁专家各自独立对检出和分类结果进行审核与修改，纠正漏诊、误诊和误判。
- 疗效评价环节：在疗效评价与分类完成之后，由于疗效评价相对简单，建议普通病例的疗效评价由1名标注医师执行，由1名审核专家进行审核。遇到复杂病历时，可酌情增加审核人数，以保证标注质量。

8 数据计算

8.1 数据计算引擎

数据计算引擎是支撑数据计算的关键，可以为感染性休克各类数据加工以及数据计算和分析任务提供高性能的算力支撑。计算引擎主要有批量计算引擎、流式计算引擎、科学计算引擎三大类。

8.1.1 批量计算引擎

主要用于大数据量、时间不敏感的数据处理和计算场景，其特点是数据吞吐量大、数据处理过程复杂、延时高；支撑感染性休克数据集大批量的数据加工、大规模数据的清洗转换以及挖掘分析等。如对批量数据清洗与转换、决策支持系统中的大数据量统计分析等业务场景。

8.1.2 流式计算引擎

主要用于对数据加工计算和应用有较强的时效性要求场景，其特点是数据处理时效高、响应速度快；用于支撑感染性休克数据中心大屏的实时动态展现、风险预警等场景。

8.1.3 科学计算引擎

科学计算引擎是一种分布式场景下大规模数据科学计算的新计算库（引擎），提升在图像处理、机器学习、深度学习等领域的计算性能，感染性休克数据集数据挖掘模型、深度学习模型的开发等应用场景需要用到科学计算引擎。

8.2 数据计算流程

对感染性休克数据的计算可遵循以下流程（图 4），也可针对感染性休克实际的医疗应用场景进行设计和改进。

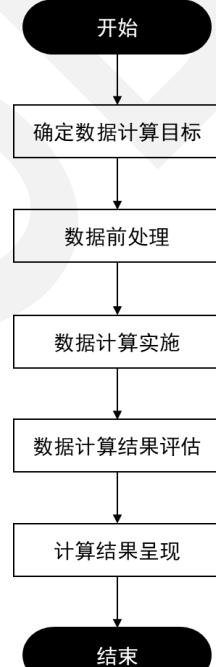


图 4 数据计算流程

8.3 确定数据计算目标

需要由临床专家和数据科学家根据感染性休克的实际医疗场景共同确立数据计算的目标，比如感染性休克患者的生存结局预测或治疗方法合理性分析等。

8.4 数据计算前处理

面向数据计算的数据前处理也包括数据去重、数据过滤、数据缺失值填补以及数据增强等（见“7.2 数据处理”部分），数据处理方法与前述大致相同。不同之处在于需要依据数据计算目标针对性地进行数据处理设计，比如在针对感染性休克病人的横断面研究和纵向研究时采取不同的数据去重处理方法等。

除上述数据处理步骤以外，通常还涉及数据采样和数据标准化等数据处理操作。

8.4.1 数据采样

由于通常不会针对数据库中的全量数据进行数据计算，所以往往会涉及到对数据进行采样操作，主要包含针对数据样本的采样、针对数据特征的采样以及时序数据的采样等等。

8.4.1.1 数据样本采样

数据样本的采样方法主要包括系统采样、随机采样及分层采样等等。

- a) 系统采样：指从总体中按照一定的标准和方法抽取一定数量的样本进行研究。优点在于简便易行，缺点在于如果总体中存在潜在的分布模式则可能引入抽样偏差。
- b) 随机采样：随机采样确保在采样过程中每个元素被采样到的几率相等，最大限度地减少了采样偏差并确保样本准确地代表了总体，是一种被普遍采用的采样方法。
- c) 分层采样：分层采样指根据样本的年龄、性别或收入水平等特征将总体进行子集划分，并在每个子集内采样一定数量的样本。该方法确保了采样到的数据样本对每个子集的代表性。

在实际数据计算或研究中，通常会联用两种或两种以上的数据采样方法构建需要进行研究的数据集，比如分时间段（如年代）对感染性休克的病例数据进行系统采样，在对照试验中根据病人是否发生感染性休克进行分层采样构建对照组和试验组，当对照组和试验组样本不平衡时在样本量较多的子集内部采用随机采样从而平衡数据集等。

8.4.1.2 数据特征采样

数据特征的采样方法包括基于经验、基于统计方法以及基于机器学习的采样方法等。

- a) 基于经验的特征采样：指由临床专家通过经验和领域知识进行特征的采样设计，比如从医学检验检查项目中选取合适的检查项数据作为需要进行计算的特征。
- b) 基于统计方法的特征采样：比如基于方差分析、相关系统及统计检验等统计学方法来进行特征

的纳入和排除。

c) 基于机器学习的特征采样：比如基于 L1 正则化、机器学习模型的特征重要性以及递归特征消除等机器学习方法进行特征的采样和选择。

8.4.1.3 时序数据采样

时序数据通常是指按时间顺序排列的数据点序列，对时序数据的采样需要考虑数据性质、数据分析和计算目标、采样频率、数据变化趋势以及计算资源开销等方面，采用均匀采样、滑动窗口采样、不规则时间采样及事件触发采样等方法进行数据采样。

8.4.2 数据标准化

数据标准化的主要目的是将不同尺度、不同范围及不同计量方式的数据转化为具有统一标准的形式，以便后续更好地进行数据计算和分析。数据标准化包括但不限于以下方面：

8.4.2.1 计量单位标准化

对以不同计量单位进行记录的数据（如千克和克、米和厘米等）用统一的计量单位进行数据转换和处理。

8.4.2.2 数值标准化

采用零均值标准化、最小-最大标准化、对数变换等方法对数值型数据进行标准化处理，使不同取值范围的数据映射到统一的区间，从而可以消除数据量纲对数据计算模型的影响、提升模型的收敛速度以及提高模型的性能和泛化能力。

8.4.2.3 图像数据标准化

图像数据的标准化处理包括图像像素值的标准化，处理方法与上述数值型标准化处理相同。除此之外，图像数据标准化还包括图像尺寸调整、图像去噪以及通道交换等，以保证图像数据计算模型的稳定性和训练效果。

8.5 数据计算实施

数据计算实施主要指根据已有的数据计算目标，选取合适的数据计算和分析方法并进行数据分析，按照分析目标层级可将数据计算实施分为描述性分析、探索性分析、因果关联分析以及预测性分析等。

8.5.1 描述性分析

描述性分析主要通过统计学方法对数据的模式、分布和基本结构进行总结和描述，主要包括：

a) 描述性统计指标计算：如计算数据的均值、中位数、标准差等基本统计指标，以了解数据的集中趋势和离散程度。

- b) 频次分析：计算数据不同取值的频率和比例，从而获取数据的分布情况。
- c) 相关性分析：计算数据不同变量之间的相关系数，从而评估变量之间的关联关系。

8.5.2 探索性分析

探索性分析旨在理解数据的结构、特征、分布、异常值以及变量之间的关系，并进行潜在的数据模式挖掘（如聚类等），从而为后续更正式的建模分析提供基础、生成假设和指导数据预处理。探索性分析的主要内容包括：

- a) 缺失和异常值识别及处理：通过识别缺失模式和潜在异常点，并决定处理策略和技术。
- b) 进一步探索数据分布形态：采用偏度、峰度、核密度估计图、分位数图等指标和图表进一步描述数据分布情况，通过检验分布假设描述数据分布特性（如正态性等）。
- c) 聚类分析：基于特征将数据进行潜在的亚类或群体划分，识别数据潜在的模式或结构，用于识别患者的疾病亚型、实现后续患者的分层以及更精准地治疗和管理。

8.5.3 因果关联分析

因果关联分析旨在确定数据中一个变量（处理/干预/原因）是否导致了另一个变量（结果/效应）的变化，核心在于识别数据变量之间的“因果效应”。因果关联分析的主要内容包括：

- a) 定义因果问题：明确干预（Treatment）、结果（Outcome）以及目标人群。
- b) 识别混杂因素：识别既影响干预分配（或暴露）又影响结果的变量。
- c) 构建因果模型：使用图形化工具（如因果图/Directed Acyclic Graphs）或潜在结果框架（counterfactual framework）来清晰地表达关于变量间因果关系的假设。
- d) 估计因果效应：在控制或调整了混杂因素后，量化干预对结果的平均效应（如平均处理效应-ATE）或特定子群体的效应（如条件平均处理效应-CATE）。
- e) 敏感性分析：评估结论对未观测混杂因素或模型假设（如可忽略性、一致性、正定性）的稳健性。

8.5.4 预测性分析

预测性分析旨在利用历史数据来预测未来的趋势或结果，主要包括：

- a) 回归分析：利用回归模型和其他变量之间的关系，进行感染性休克病情发展、药物疗效等预测。
- b) 分类分析：将数据进行有监督的分类标签预测，以识别患者疾病类别标签等，有助于疾病早期干预、个性化治疗，提高诊断准确率和效率。
- c) 时间序列分析：分析数据的趋势、季节性和周期性，为未来的预测提供依据。

8.6 计算结果呈现

对数据计算结果的呈现包括但不限于结果可视化及结果报告撰写等。

8.6.1 结果可视化

使用图表、地图及视频等方式将数据计算结果呈现出来，以直观地传达数据计算所得到的信息和趋势。

8.6.2 结果报告撰写

将数据计算的结果整理成报告或演示文稿，并附加结论性的文字描述，以这种清晰简洁的方式传达数据计算结果。

9 互操作性

9.1 数据脱敏互操作

针对电子病历数据，支持病人姓名、年龄、性别、ID号、检查日期、出入院日期的脱敏。针对医学影像数据，支持病人姓名标识的脱敏。

用户上述数据，点击“数据脱敏按钮”，经平台自动计算后，回显脱敏后数据结果，并支持用户进行下载。其中，敏感信息均用“****”表示。

9.2 数据处理互操作

9.2.1 数据清洗操作

支持电子病历数据和医学影像数据的清洗操作。具体地，针对电子病历数据，支持非空检核、主键重复检核、非法代码清洗、非法值清洗、数据格式检核、记录数检核及业务约束核验操作。针对医学影像数据，支持检查数据扫描质量及检查数据扫描完整性操作。

用户上传数据，点击“数据清洗按钮”，经平台自动计算后，回显清洗后数据结果，并支持用户进行下载。

9.2.2 脏数据处理操作

支持电子病历数据和医学影像数据的脏数据处理。具体地，针对电子病历数据，支持处理空缺值、离群值和不一致等脏数据。针对医学影像数据，支持处理缺失值数据、错误数据和错误关联数据。

用户上传数据，点击“脏数据处理按钮”，经平台自动计算后，回显处理后数据结果，并支持用户进行下载。

9.3 数据计算互操作

平台依托于批量计算引擎、流式计算引擎及科学计算引擎，进行数据计算和分析操作。

9.3.1 数据采样互操作

支持针对数据样本、针对数据特征以及时序数据的采样，共计3种采样方式。不同采样策略的具体采用参数需要由用户手动录入并上传。具体操作为：

- a) 用户上传原始数据，选择所需的采样方式，并输入对应的采样参数；
- b) 平台接收信息后，自动执行采样计算；
- c) 计算完成后，用户可下载采样结果。

9.3.2 数据标准互操作

针对电子病历数据，支持计量单位标准化和数值标准化。针对医学影像数据，还支持面向图像的特定标准化策略。具体操作为：

- a) 用户上传原始数据，选择标准化方式；
- b) 平台接收信息后，自动执行标准化计算；
- c) 计算完成后，用户可下载标准化结果。

9.3.3 数据分析互操作

针对原始数据统计性描述的需求，支持描述性统计指标计算、频次分析及相关性分析操作。针对未来趋势预测的需求，支持回归分析、分类和聚类分析及时间序列分析操作。此外，支持机器学习方法，用于关联规则挖掘。具体操作为：

- a) 用户上传原始数据，选择所需的分析方式，并输入对应的算法参数；
- b) 平台接收信息后，自动执行分析计算；
- c) 计算完成后，用户可下载分析结果。

10 数据安全

10.1 数据访问权限控制

感染性休克数据集采用多级授权管理模式，对用户权限进行管理，可针对用户、数据等进行权限配置。主要包括授权权限管理和业务授权管理。

- a) 授权权限管理：中心系统管理员给业务管理员授权，授权业务管理员对其业务范围内的操作授权。
- b) 业务授权管理：各业务管理员根据数据集管理员所授业务范围的权限，对业务范围内的具体业务操作权限授权。

10.2 数据加密

数据加密分为存储加密和传输加密。

a) 存储加密：即感染性休克数据集的数据先加密再存储，防止重要数据丢失，防止数据文件被复制、篡改或盗用；加密过程中选择加密强度较高的加密算法，提供数据逻辑性和有效性的自动校验功能，对用户输入信息进行安全检查，降低 SQL 注入，数据库管理员权限被篡改、滥用等数据安全风险。

b) 传输加密：采用加密技术实现用户身份和鉴权口令、用户资源等关键信息的加密传输或加密存储，防止信息在网络传输或存储中被窃取、破坏及篡改。数据加密可支持多种加密形式及数据加密提醒。数据存储加密一定程度上影响运算及读取速度，可按需调整。

10.3 安全审计

感染性休克数据集使用应该采取以下审计措施：

- a) 审计内容宜包括人员审计、管理审计、技术审计（系统、网络、操作、日志审计等）；
- b) 任何操作，包括登录、创建、修改和删除记录的行为，都宜自动生成带有时间标记的审计记录，包括但不限于修改时间、修改原因、修改内容、修改人及签名等信息，并可供审计；
- c) 应制定和部署感染性休克数据集活动审计政策，重点对于感染性休克数据集的访问及操作的合规性进行审计，确定必要的审计控制范围和需要审计的数据；
- d) 应制定适当的标准操作流程，确定异常报告所需的审计跟踪数据和监视程序的类型；
- e) 审计记录应安全存储并实施访问控制，只允许授权人员能够查看相关记录，保存的内容需反映临床医学研究整个过程。

10.4 数据操作日志

感染性休克数据集应该建立数据访问操作日志记录，通过高易用性的日志功能，对所有应用系统的操作过程轨迹进行记录，以便为以后对操作痕迹进行追踪，为应用系统操作漏洞分析提供原始证据。访问日志记录，包括了操作轨迹的记录、应用系统错误日志记录、访问日志数据存储和分析等。以访问日志的形式，确保应用操作的不可抵赖性。

10.5 数据备份

本地备份是数据库容灾的重要组成部分，需要建立健全的备份和恢复机制，对感染性休克数据集/库进行保护。通过备份恢复设备，对重要数据进行备份保护。根据业务特点及数据性质，制定相应的备份策略，并安排运维人员，对备份进行监控，如遇备份失败，第一时间进行故障定位并重启备份任务。建议备份策略：

- a) 根据数据库厂商官方的备份建议，以周为备份周期；
- b) 周日全备份；
- c) 周一至周六增量备份；

- d) 每天每 12 小时执行一次备份（依据实际日志产生数量）；
- e) 备份保留周期，依据实际要求确定。

11 数据共享

数据共享应符合国家、地区等各级法规及制度规定。医疗大数据的统一汇聚共享和治理为医疗和科研提供了支撑，数据在经过严格的数据清洗与质量把控后，满足内外资源整合共享的基本要求。依托云平台的计算和存储能力，基于“数据中台”和“业务中台”双中台为核心的大数据平台实现医疗数据共享功能。

具体来说，数据中台负责整合医院的医疗健康数据，根据数据特点和应用需求，构建分布式数据仓库，高效分类存储结构化和非结构化数据。具体来说，海量医疗数据通过数据元数据理念与技术形成统一存储的“数据湖”，随后对“数据湖”中的数据资源进行整合归一，并集中存储形成分门别类的“数据仓库”并进行统一管理和维护。通过“湖仓一体”的数据存储管理模式，增强数据治理能力，确保了数据的规范性和可信度，进一步降低数据抽取和调用两个过程中的耦合性；业务中台负责整合医联体之间的业务，根据业务协同应用需求，构建统一的业务协作平台，形成“数据共享+协同服务”的运行模式，实现多院区、医联体成员单位、产业机构之间的业务协同。

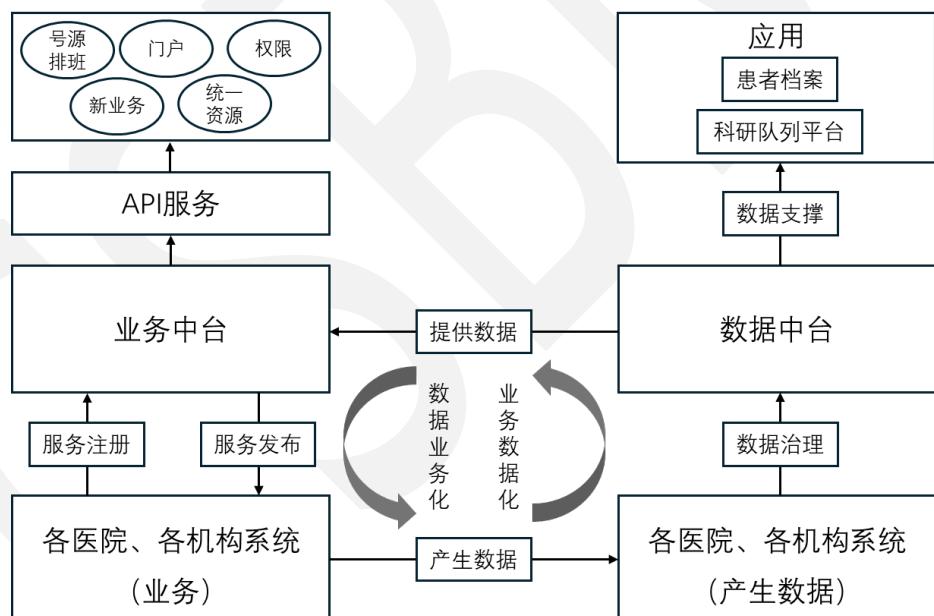


图 5 基于双中台的大数据平台流程示例

附录 A

(资料性附录)

感染性休克队列标准数据集

根据感染性休克数据采集、标注和质量控制等标准，本示例为建立感染性休克队列数据集的规范示例。本数据集以 18 岁以上成人病例为主要研究对象，数据规范以全周期病历为主，床旁实时生命体征为辅，面向感染性休克相关临床药械科研、人工智能辅助诊疗的研发和测试等需求，指定病例数据的标准化采集、标注和质控等标准，并建设感染性休克标准数据库。标准数据库的数据元字段如下：

A.1 基本信息

基本信息	
出生日期 (年/月/日):	_____
年龄:	_____ 岁 (自动计算)
性别:	<input type="checkbox"/> 男 <input type="checkbox"/> 女
民族:	<input type="checkbox"/> 汉族 <input type="checkbox"/> 其他: _____
身高 (cm):	_____
体重 (Kg):	_____
住院号:	_____
生命体征: 心率 (次/分):	_____
体温 (°C):	_____
血压 (mmHg):	_____
婚姻状况:	<input type="checkbox"/> 已婚 <input type="checkbox"/> 离异 <input type="checkbox"/> 丧偶 <input type="checkbox"/> 单身
宗教信仰:	<input type="checkbox"/> 天主教 <input type="checkbox"/> 基督教 <input type="checkbox"/> 佛教 <input type="checkbox"/> 无 <input type="checkbox"/> 其他: _____
工作状况:	<input type="checkbox"/> 全职 <input type="checkbox"/> 兼职 <input type="checkbox"/> 家庭主妇 (夫) <input type="checkbox"/> 学生 <input type="checkbox"/> 退休 <input type="checkbox"/> 无业

A.2 现病史

A.2.1 主诉

主诉	
主诉内容:	_____。
诊断方式:	<input type="checkbox"/> 临床诊断 <input type="checkbox"/> 影像诊断 <input type="checkbox"/> 未知

A.3 既往史

A.3.1 饮酒及吸烟史

饮酒史	
-----	--

是否有饮酒史: 从未 已戒酒 仍饮酒

饮酒类型: 白酒 红酒 啤酒

若“已戒酒”或“仍饮酒”，持续时长（月）:

平均每日饮酒总量 (ml):

若“已戒酒”，最近一次戒酒日期（年/月/日）:

吸烟史

是否有吸烟史: 从未 已戒烟 仍吸烟

若“已戒烟”或“仍吸烟”，持续时长（月）:

平均每日吸烟数量（支）:

若“已戒烟”，最近一次戒烟日期（年/月/日）:

A.3.2 既往病史

既往病史

是否存在既往病史: 是 否（若是，填写以下内容）

疾病名称	治疗方式	发病日期（年/月/日）	是否持续	是否为精神疾病
		<u> </u>	<input type="checkbox"/> 是 <input type="checkbox"/> 否	<input type="checkbox"/> 是 <input type="checkbox"/> 否
		<u> </u>	<input type="checkbox"/> 是 <input type="checkbox"/> 否	<input type="checkbox"/> 是 <input type="checkbox"/> 否
		<u> </u>	<input type="checkbox"/> 是 <input type="checkbox"/> 否	<input type="checkbox"/> 是 <input type="checkbox"/> 否
		<u> </u>	<input type="checkbox"/> 是 <input type="checkbox"/> 否	<input type="checkbox"/> 是 <input type="checkbox"/> 否
		<u> </u>	<input type="checkbox"/> 是 <input type="checkbox"/> 否	<input type="checkbox"/> 是 <input type="checkbox"/> 否

A.4 检验检查结果

A.4.1 体格检查

本次入院体格检查			
部位	未查	结果	异常描述
一般情况	<input type="checkbox"/>		
头	<input type="checkbox"/>		
眼、耳、鼻、喉	<input type="checkbox"/>		
颈	<input type="checkbox"/>		
心脏	<input type="checkbox"/>		
胸部	<input type="checkbox"/>		
腹部	<input type="checkbox"/>		
四肢	<input type="checkbox"/>		
皮肤	<input type="checkbox"/>		
淋巴结	<input type="checkbox"/>		
神经系统	<input type="checkbox"/>		

A.4.2 实验室检查结果

实验室检查					
项目分类	项目检查	检查结果	单位	参考值	临床评估
血常规	样本采集日期 (年/月/日) : <input type="text"/> - <input type="text"/> - <input type="text"/>				
	红细胞计数		$10^{12}/L$		
	血红蛋白		g/L		
	红细胞压积		%		
	平均红细胞体积		fL		
	平均红细胞 HGB 含量		Pg		
	平均红细胞 HGB 浓度		g/L		
	RBC 分布宽度 SD		fL		
	RBC 分布宽度 CV		%		
	血小板计数		$10^{9}/L$		

白细胞	白细胞计数		10 ⁹ /L		
	中性分叶核粒细胞百分率		%		
	淋巴细胞百分率		%		
	单核细胞百分率		%		
	嗜酸性粒细胞百分率		%		
	嗜碱性粒细胞百分率		%		
	原始细胞百分率		%		
	中性分叶核粒细胞绝对值		10 ⁹ /L		
	淋巴细胞绝对值		10 ⁹ /L		
	单核细胞绝对值		10 ⁹ /L		
血生化	嗜酸细胞绝对值		10 ⁹ /L		
	嗜碱细胞绝对值		10 ⁹ /L		
	样本采集日期 (年/月/日) :	_____	_____	_____	_____
	谷内转氨酶 (ALT)		U/L		
	天门冬氨酸氨基转移酶 (AST)		U/L		
	总胆红素 (TBIL)		umol/L		
	直接胆红素 (DBIL)		umol/L		
	间接胆红素		umol/L		
	丙氨酸氨基转移酶		IU/L		
	门冬氨酸氨基转移酶		IU/L		
生化	总蛋白		g/L		
	白蛋白		g/L		

球蛋白		g/L		
白球比例				
葡萄糖		mmol/L		
尿素		mmol/L		
肌酐		umol/L		
估算肾小球滤过率				
血清胱抑素 C 测定		mg/L		
尿酸		umol/L		
甘油三酯		mmol/L		
胆固醇		mmol/L		
高密度脂蛋白		mmol/L		
低密度脂蛋白		mmol/L		
碱性磷酸酶		IU/L		
谷氨酰转肽酶		IU/L		
肌酸激酶		IU/L		
乳酸脱氢酶		IU/L		
羟丁酸脱氢酶		IU/L		
血清钾 (K)		mmol/L		
血清钠 (Na)		mmol/L		
血清氯 (Cl)		mmol/L		
血清钙 (Ca)		mmol/L		
血清磷 (P)		mmol/L		

	血清镁 (Mg)		mmol/L		
	二氧化碳结合力		mmol/L		
	阴离子间隙		mmol/L		
	血清 β 羟基丁酸测定		mmol/L		
	血清无机磷		mmol/L		
	总胆汁酸		umol/L		

临床评估*: 1=正常; 2=异常, 但无临床意义; 3=异常, 并有临床意义; 4=未进行检查。若选择“异常有临床意义”, 请确认是否需要记录既往病史或不良事件。

A.4.3 床旁实时生命体征检查

床旁实时生命体征监测	
是否进行神经电生理监测: <input type="checkbox"/> 是 <input type="checkbox"/> 否 (若是, 填写以下内容)	
监测时间 (年/月/日): _____	
检查参数:	

A.4.4 其他检查评估

其他检查评估	
是否进行其他检查评估: <input type="checkbox"/> 是 <input type="checkbox"/> 否 (若是, 填写以下内容)	

A.5 治疗情况

A.5.1 抗感染药物治疗

抗感染药物治疗						
是否接受过抗感染药物治疗: <input type="checkbox"/> 是 <input type="checkbox"/> 否 (若是, 填写以下内容)						
治疗	药物	使用	用药时间	进展或复发日期 (年/月/日)	停止治	其他

类型	名称	途径			疗原因	
			□□□□-□□-□□	□□□□-□□-□□		
			□□□□-□□-□□	□□□□-□□-□□		
			□□□□-□□-□□	□□□□-□□-□□		
			□□□□-□□-□□	□□□□-□□-□□		
			□□□□-□□-□□	□□□□-□□-□□		

治疗类型：

药物名称（通用名）：

使用途径： 1=口服； 2=静脉； 3=肌肉注射； 4=皮下注射； 5=皮内注射； 6=吸入； 7=雾化吸入； 8=外用； 9=局部使用； 其他，请直接填写。

停止治疗原因： 1=疾病进展； 2=完成规定的治疗； 3=毒性不耐受； 4=患者决定； 99=其他。

A.6 疗效评价

疗效评价				
治疗方式*	治疗方式详述	开始时间（年/月/日）	结束时间（年/月/日）	最佳疗效*
		□□□□-□□-□□	□□□□-□□-□□	
		□□□□-□□-□□	□□□□-□□-□□	
		□□□□-□□-□□	□□□□-□□-□□	
		□□□□-□□-□□	□□□□-□□-□□	
		□□□□-□□-□□	□□□□-□□-□□	

最佳疗效*： 1=完全缓解； 2=部分缓解； 3=疾病稳定； 4=疾病进展； 5=无法评价； 6=未知