**“全新呼吸”中国哮喘患者在线登记**

**数据库系统II期**

**数据管理计划**

|  |  |  |  |
| --- | --- | --- | --- |
| 版本号 | 编制 | 审核 | 日期 |
| V# |  |  |  |
| V# |  |  |  |
| V# |  |  |  |

**目录**

[1. 编制说明 4](#_Toc113456197)

[2. 数据管理计划目标 4](#_Toc113456198)

[3. 相关人员 4](#_Toc113456199)

[4. 启动阶段 5](#_Toc113456200)

[数据库设计 5](#_Toc113456201)

[数据库测试 5](#_Toc113456202)

[数据安全管理 6](#_Toc113456203)

[5. 运行阶段 7](#_Toc113456204)

[数据录入 7](#_Toc113456205)

[数据核查与质疑 7](#_Toc113456206)

[医学编码 8](#_Toc113456207)

[数据质量控制 9](#_Toc113456208)

[期中分析 9](#_Toc113456209)

[6. 结束阶段 9](#_Toc113456210)

[数据库的锁定 9](#_Toc113456211)

[数据库的解锁 10](#_Toc113456212)

[数据和数据管理文件的归档 10](#_Toc113456213)

**缩写说明（待完善）**

# 编制说明

为确保“全新呼吸-中国哮喘患者在线登记数据库系统建设项目II期”项目的数据质量，最大程度地保证登记数据库的安全和实现登记数据库的效用，提前制定计划规范数据的采集、管理、储存、清洗和分析等过程。验数据管理，

# 数据管理计划目标

本数据管理计划旨在规范数据管理过程，以保证数据的真实性、完整性、一致性、准确性。

# 相关方参与计划

|  |  |  |
| --- | --- | --- |
| 角色 | 任务 |  |
| 总中心项目组 | 制定CRF、质控 |  |
| 云平台公司 | 数据采集平台搭建、测试、培训、答疑 |  |
| 数据录入员 | 数据录入 |  |
| 南鹏PM | 质疑、溯源 |  |
| 南鹏数据管理员 | 逻辑核查、质疑、数据质量报告 |  |
| 南鹏统计分析师 | 统计分析 |  |

# 启动阶段

## 4.1 数据库设计

本登记研究利用中国哮喘患者在线登记云平台系统（下文简称为登记云平台）进行CRF表单中的数据采集。

登记云平台的设计应包括以下两个要素：1)根据CRF表决定采集的数据要素，以及根据注释CRF确定数据要素的变量名称、类型、格式、长度、编码赋值、其他描述；2）建立逻辑核查需求（edit check specification, ECS），设计数据要素的逻辑检查规则，以及不同数据元素之间的逻辑关系。

## 4.2 数据库测试

数据库的测试主要包含对录入、导入、导出、逻辑核查功能的测试。

|  |  |
| --- | --- |
| **测试目的** | **测试内容** |
| 录入 | 测试数据是否能录入、是否能完整录入、是否缺少录入数据点，测试是否能对“异常”数据提出质疑 |
| 导入/导出 | 测试外部电子化数据的加载是否正确，测试导出数据与导入数据的一致性，如数值、类型、格式、变量名称和标签、导出数据的完整性和导出逻辑 |
| 权限 | 确保用户可以实现权限范围内所有能并不能实现其他角色权限功能 |
| 逻辑核查 | 测试数据错误情况，如数据缺失、正常值范围之外的数据、数据类型错误等，是否能提出数据质疑 |

测试人员安排：

录入测试-总中心项目组

导入导出-南鹏研究院数据管理部门

权限-总中心项目组、参与单位、PM

逻辑核查-总中心项目组

## 数据安全管理

应在首例受试者入组前完成数据管理系统的安全性验证。对系统中不同职责的使用者授予不同的权限。各项目成员的权限如下：

|  |  |  |
| --- | --- | --- |
| 人员 | 职责 | 权限 |
| CRA | 原始数据核查 | 所负责中心数据的核查，发送质疑及关闭本人质疑 |
| 各医院CRC | 录入 | 本中心数据录入 |
| DM | 数据管理 | 项目所有数据逻辑核查，发送质疑，关闭质疑 |
| Medical | 医学经理 | 项目所有数据的医学核查，发送质疑，关闭本人质疑 |
| PM | 项目经理 | 查看所有中心数据、导出数据 |

数据备份

云平台公司应对数据进行定期备份，可每日在其他云平台对数据进行备份更新。

# 运行阶段

## 数据录入

数据录入员在经过培训后，利用登记云平台授权的用户信息进行数据的录入和修正。

## 数据核查与质疑

主要从数据的**真实性、完整性、准确性、时效性和可溯源性**进行数据核查。按照逻辑核查所涉及字段的多少，将从**单变量**和**多变量**分别进行核查：

1. 单变量核查指每次只检查一个字段，常见于以下情况：

|  |  |
| --- | --- |
| 问题分型 | 案例 |
| 完整性 | 必填数据缺失 |
| 准确性 | 数据录入错误  例：实验室检查、年龄等超出规定或合理的范围，录入未来的日期 |
| 可溯源性 | 数据与原始病历记录不符合 |
| 数据标准化 | 数据格式不符（数值格式、数值长度、日期格式、文本长度等）  例：数字型内录入文本，日期内录入非日期格式 |

1. 多变量核查指每次核查**两个或者两个以上的字段**前后矛盾或者逻辑是否合理，如男性被勾选妊娠检查。在进行多变量核查前，**必须首先完成**数据的单变量核查。
2. （举例如下：）
3. 

如果题3不包含题2的选项，则逻辑错误。



13题=‘有’，与12题=‘无急性发作’同时存在，则逻辑错误

本登记研究中的核查要点**\***包括：

* 一般资料
* 病史资料
* 既往病史
* 卫生支出
* 实验室检查
* 问卷评估
* 疾病诊断
* 严重程度评估

参照以上核查要点，如发现数据存在缺失、不完整、不合理、不准确、数据间或数据和源数据不一致等，则需要在登记系统或者通过纸质数据澄清表（data clarification form, DCF）发出数据质疑。 DM则需要在患者入组首例后每个月对数据进行核查并出具《数据质量报告》。

## 医学编码

将登记云平台中收集的医学诊断、合并用药、既往用药、既往病史等描述与标准词典中的属于进行匹配。编码流程大致如下：

|  |  |
| --- | --- |
| 流程名称 | 具体内容 |
| 编码启动会议 | * 选择编码系统； * 确定词典版本； * 确定编码对象，如既往病史、合并用药； * 编辑编码指南 |
| 电子编码系统等等建立与测试 | * 确定编码对象与编码所需信息能自动导入登记数据库； * 确定编码对象能匹配到正确的词典； * 确定词典在系统中正确呈现； * 确定由编码系统发出或关闭的质疑能在登记系统中正确体现 |
| 进行医学编码 | * 自动编码； * 手动编码 |
| 审查编码报告与最终数据提交 |  |

## 数据质量控制

数据质控将对数据录入进度、完整性、一致性和可溯源性等进行质控。数据质控需要明确核查项目、核查的关键指标和非关键指标，并提供数据质量核查报告，对数据录入的完整性、错误率等进行汇报。

根据随访计划的安排，本登记研究数据质量控制时间安排如下：计划在患者入组一个月后，每个季度，各中心最后一例患者结题锁库前，分别进行数据质控。

将采用远程和现场两种方式进行质控。其中远程质控每次质控比例为10~20% （若低于20例，则全部检查）；现场质控比例为5%~10%（若低于10例，则全部检查）

## 期中分析

将在＃（ｗｈｅｎ）进行一次期中分析。分析每次数据采集和核查的范围，出具《数据质量报告》，并对数据进行描述性分析，出具《描述性分析报告》

**外部数据传输协议**

针对项目与内外部统计分析师以及客户商定进行数据传输的规范性文件。

文件命名方式及数据类别

传输方式及频率

传输内容及格式

在本文件定稿前，数据管理员应通过测试数据与统计师进行数据传输测试，并与客户确认，出现问题及时更新数据传输协议，直至测试成功，并得到客户书面确认。

# 结束阶段

## 数据库的锁定

数据库最终锁定前，应确认以下内容：

* + - 1. 方案所列所有数据已经收到并准确地录入数据库；
      2. 其他外部数据已经导入到数据库中；
      3. 已完成数据的逻辑性和一致性验证结果审查；
      4. 所有数据质疑已解答并进入数据库；
      5. 已完成医学编码；
      6. 已完成数据质量核查；
      7. 更新并保存了所有相关文档；

完成以上步骤后，可由研究相关人员（研究者、数据管理人员、生物统计师、监查员等）签字及签署日期，书面批准数据库锁定。

## 数据库的解锁

如数据库需要进行解锁，则需要详细说明数据库解锁原因、解锁审批人及再锁定的要求和执行标准文件。

## 数据和数据管理文件的归档

按照法规要求，对各个管理文件进行存档。