**自闭症等专病队列数据处理服务项目需求参数**

一、项目名称：自闭症等专病队列数据处理服务采购

二、项目背景及概述

基于出生队列随访数字化集成平台，完善自闭症、儿童肥胖等常见儿童发育疾病单病种数据库，对随访问卷、体征检查、实验室检查、儿童相关疾病信息等进行数据抄录、数据质控、数据补录、数据审查和校验，形成集检测结果、随访信息及病史等多模态数据，便于数据分析利用及科研成果产出。

三、项目总体目标

针对专病队列包括但不限于：自闭症、肥胖儿童，对患者原始数据进行全面标准化、结构化收集归纳整理；通过数据清洗服务将重复、多余的数据进行筛选和清除处理，纠正或删除错误数据，对数据各质控维度进行全面数据质控；为专病数据库提供真实、有效的数据支撑，提升专病数据价值及挖掘效率，为形成集检测结果、随访信息及病史等多模态数据做数据基石。

四、具体服务内容及要求

**1、数据服务内容**

|  |  |
| --- | --- |
| **服务内容** | **说明** |
| 数据采集服务 | 原始数据的人工采集服务 | 对自闭症、肥胖儿童的原始数据的人工采集（含病史摘抄），包括随访问卷、体征及实验室检查、儿童相关疾病信息等进行人工抄录，保证数据完整、真实、安全。共需采集2500份。 |
| 数据处理服务 | 数据质控 | 对于进入平台数据（随访问卷、体征检查、实验室检查、儿童相关疾病信息），每月进行质控，包括：数据录入二次校验、逻辑检查、缺失情况分析等，定期发布质控报告。质控数据至少包含：检测数据质控、随访问卷和临床检查数据(体征检查+化验)质控、病史摘抄数据质控。 |
| 数据补录 | 根据每月质控结果，自主设计完成数据预处理和补录方案，对质量欠佳数据补录，保证数据完整及准确性，总数量500份。 |
| 数据清洗 | 自主设计完成数据清洗方案，制定数据清洗步骤，针对全量数据审查校验，最终形成可直接用于数据分析的完整数据库。 |

**2、数据质控要求**

**2.1、数据分析**

在数据质控过程中本着纠错的原则，对元数据存在的异常值、重复值、错误值、离群值等不符合标准的原始数据予以分析，部分需要人工确认的数据，导出经业务部门确认后再对数据进行修正。

做好数据质量管理及安全管理，建立数据合理性验证、数据更正等机制。在项目实施过程中应严格按照国家信息安全保密规定，做好数据传输、质控分析过程的数据保密工作，确保数据的信息安全，不得以任何形式向第三方提供任何相关数据信息。若造成数据泄露，采购人有权追究供应商法律责任。

数据质控分析规则**至少**包括非空检核分析、主键重复分析、非法代码分析、非法值分析、数据格式检核、记录数检核。

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。

主键重复：多个业务系统中同类数据经过质控分析后，在统一保存时，为保证主键唯一性，需进行检核工作。

非法代码、非法值质控分析：非法代码问题包括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,需根据具体情况进行校核及修正。

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，**至少包含**时间格式、币种格式、多余字符、乱码。

**2.2、数据质控报告**

问题数据分析：对于质控发现的异常数据首先出具问题数据分析报告，异常数据排查完成后，将相关异常数据提供给业务科室进行分析总结并编写问题数据分析报告，制定数据清洗处理方案。

**3、数据清洗要求**

**3.1、数据清洗目标**

（1）确保数据的真实有效、清洗的最终数据是源于各业务部门的实际数据。

（2）对各分散数据进行合理、高效的采集整合。

（3）保证数据的质量、完整性、准确性。

（4）严格按照约定的格式进行数据分析、清理。

**3.2、数据清洗范围**

本次数据清洗的范围需按照数据清洗规范，对实际数据进行分析、清洗。针对全量数据审查校验（**至少**包含：随访问卷、体征及实验室检查、儿童相关疾病信息），最终形成可直接用于数据分析的完整多模态数据库。本着实事求是的前提，对数据进行归纳清洗。

**3.3、数据清洗步骤**

数据清洗工作必须包括以下过程：现有数据情况分析、清洗测试编码、数据清洗和转换（含缺失数据补录确认）。

数据情况分析：了解原数据结构、代码表、数据流程等。

清洗测试编码：根据数据导入字典对照报告、依据相关业务规则和逻辑检查编写数据清洗程序，并对数据清洗程序进行功能测试。

数据清洗、转换：用编写好的数据清洗程序对现有的数据进行清洗，对要求字段数据缺失的协助业务部门进行补录。

**3.4、数据清洗原则**

**3.4.1、数据清洗过程，应确保数据的正确性。**

（1）本着纠错的原则，对原数据存在错误的部分予以纠正。

（2）对于垃圾数据应予以清除。

（3）本着纠错的原则，对于不满足业务规则和逻辑的数据，部分需经过人工确认的数据，需要导出确认，然后纠正错误数据。

**3.4.2、数据清洗过程，应确保数据的延续性。**

（1）对原数据在逻辑关系上不符合业务规则和逻辑要求的，对业务数据进行清洗，以保证数据清洗的合理性。

（2）对于代码信息项（数据编码），参照出生队列随访数字化集成平台的数据编码进行相应对照。

**4、数据预处理、补录要求**

基于质控分析出数据问题产生的原因，对问题数据进行归类处理：对于可以通过信息化进行批量数据处理的情况，由业务科室确认处理规则后，编写数据清理脚本进行数据处理。主要可以处理的内容**至少**包括：数据格式类错误修正、业务表中间数据遗漏数据修正、经确认后可以按照统一规则进行的数据错误修正、对于有规律的错误数据由业务人员按照统一格式整理成数据文件导入系统后统一修正。

对于需要由业务人员通过业务系统的前台操作完成数据修正的情况，由业务人员根据档案信息或者联系服务对象重新收集数据。

**5、质控及治理报告要求**

基于全量数据、编制数据分析质控报告，综合分析呈现数据的现状、问题、特点，以及问题分析改进和清洗方案，出具数据质控分析报告。

根据数据质控分析报告，对数据进行清洗，数据清洗的情况，编写数据清洗治理报告。最终给出符合采购人要求的规范数据。

**6、服务质量的审核要求**

供应商质控分析报告需自检合格后方可递交采购人审核。

采购人对供应商筛选清理、预处理、质控分析及整合的报告进行审核。审核以抽检方式进行，根据采购人提供的数据标准进行检查，要求数据准确率达到 100%,予以审核通过。抽检合格率低于100% 的，提交审核的数据全部发回供应商重新自检，自检合格后再提交审核。经三次审核仍不合格的，扣除本项目合同金额5%并再次全部发回供应商重新自检，自检合格后再提交审核，此后的每次不合格再扣除合同金额的 5%，最高扣除金额为合同全款的50%。

五、售后服务要求

|  |  |  |
| --- | --- | --- |
| **序号** | **项目内容** | **要求** |
| 1 | **服务期限** | 合同签订后3个月内。 |
| 2 | 服务响应 | 在接到用户通知后，2小时内响应，4小时内到达现场，超过12小时不能解决问题应提供应急方案。 |
| 6 | 售后服务 | 供应商应承诺保质保量地完成相关数据服务。通过线上或线下的方式，提供**不少于1年**的免费维护和技术支持，时间从验收合格之日起计算。供应商应承诺超过免费维护、升级和服务时间后，年服务费用的收费标准。 |

六、项目实施要求

1、供应商应具备本地化服务能力，在本地设有售后服务机构及项目支持团队（提供售后服务机构地址证明及项目支撑人员社保证明），能够快速响应采购人的各种使用问题及故障处理。为使本项目按质、按量、按时及有序实施，供应商必须有一个完善和稳定的管理组织机构。

2、在项目负责人员方面，供应商须配备一名专职负责的项目负责人。在服务期内，质控分析实施团队必须在采购人指定的地点开展质控分析工作。**要求配备具有数据分析能力和经验的人员。**

3、本项目的项目负责人、质控分析实施人员必须对质控分析的数据有足够的认识。供应商如中途更换项目负责人、技术负责人，则必须征得采购人同意。

4、在项目实施期间和质保期内，供应商承诺的项目负责人和开发实施的主要人员未经采购人同意不得调整。

5、项目实施期间，供应商必须设置专人在项目实施期间对文档进行检查和管理，项目最终验收后全部移交采购人。

6、供应商须对项目实施过程中涉及的相关数据承担保密义务。

七、付款方式

采购人在项目验收合格后三个月内付清全款。