# 基于REDCap系统开发大型自然人群队列数据质控平台及其实践使用说明

**1.系统概况**

**1.1目的和意义**

医学领域中大部分研究都需要收集大量数据，并依赖于对所收集数据的分析而展开。研究过程中的数据质量控制非常重要，决定了研究的质量和成败。以往，医学研究常采用纸质问卷收集数据，并应用EpiData软件进行数据双录入和质量控制。EpiData软件录入快捷、功能完备，是医学研究中广泛采用的数据管理和质量控制工具，但该软件为单机版程序，仅可进行单中心数据质量控制，没有联网核查功能，不适用于多中心研究对数据质量控制工作发展的要求。越来越多的医学研究采用电子数据收集系统（Electronic Data Capture，EDC）收集研究数据，以保证数据质量和过程管理的规范。但商业EDC价格高昂、配置复杂、学习和使用成本高，难以在研究（尤其是探索性研究）中广泛应用。广大临床医生及科研人员急需一种低成本、方便易用的电子数据收集系统。REDCap（Research Electronic Data Capture）就是满足这一要求的工具，该系统由范德堡大学的Paul Harris教授团队研发，具有易上手、便捷、安全、稳定等多种优点，近年来在国内外医学研究中得到广泛应用。REDCap系统数据录入方式多样，既可在已有纸质问卷的基础上，进行数据的电子化录入及质量控制，也可开放给调查对象直接在线填写一次问卷，再由研究人员对问卷进行审核。且该系统应用范围广，通过不同的功能模块设置可实现单次研究数据、长期纵向数据等多种类型数据的录入。目前医学研究中，收集纸质问卷后再进行数据电子化录入的方式应用较多，考虑到数据质量控制在医学研究中的重要性，本文将依托西北自然人群队列研究数据库建立过程中的实践经验，针对性地介绍在已有纸质问卷基础上，利用REDCap系统实现研究数据质量控制的方法。

**1.2 REDCap的通用流程**

REDCap系统主要应用于临床研究电子化数据采集与数据库构建管理，该系统的开发始于2004年，主要应用于临床研究，经过十余年的完善，现已被广泛应用于多个研究类型，包括横断面调查、临床试验、队列研究等，内容涵盖临床实践质量评估、效果评价、基础科学研究、患者调查、临床决策等。REDCap的界面交互友好、操作直观且可重复使用，依托REDCap系统进行大型自然人群队列研究的通用流程图见图（1-1）。

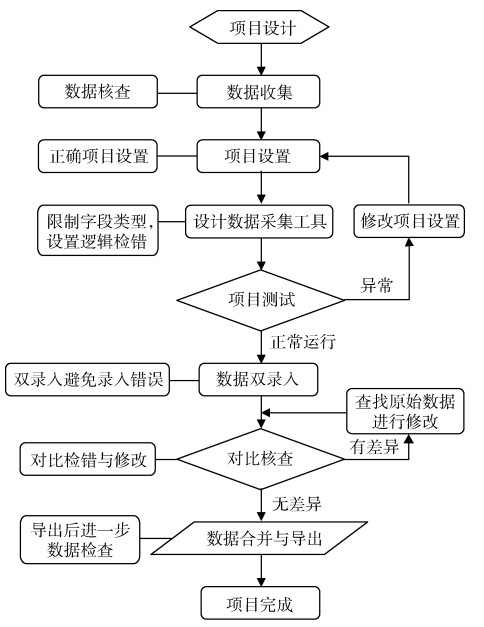


图 1-1 REDCap进行大型自然人群队列研究流程图

**1.3项目主界面**

****

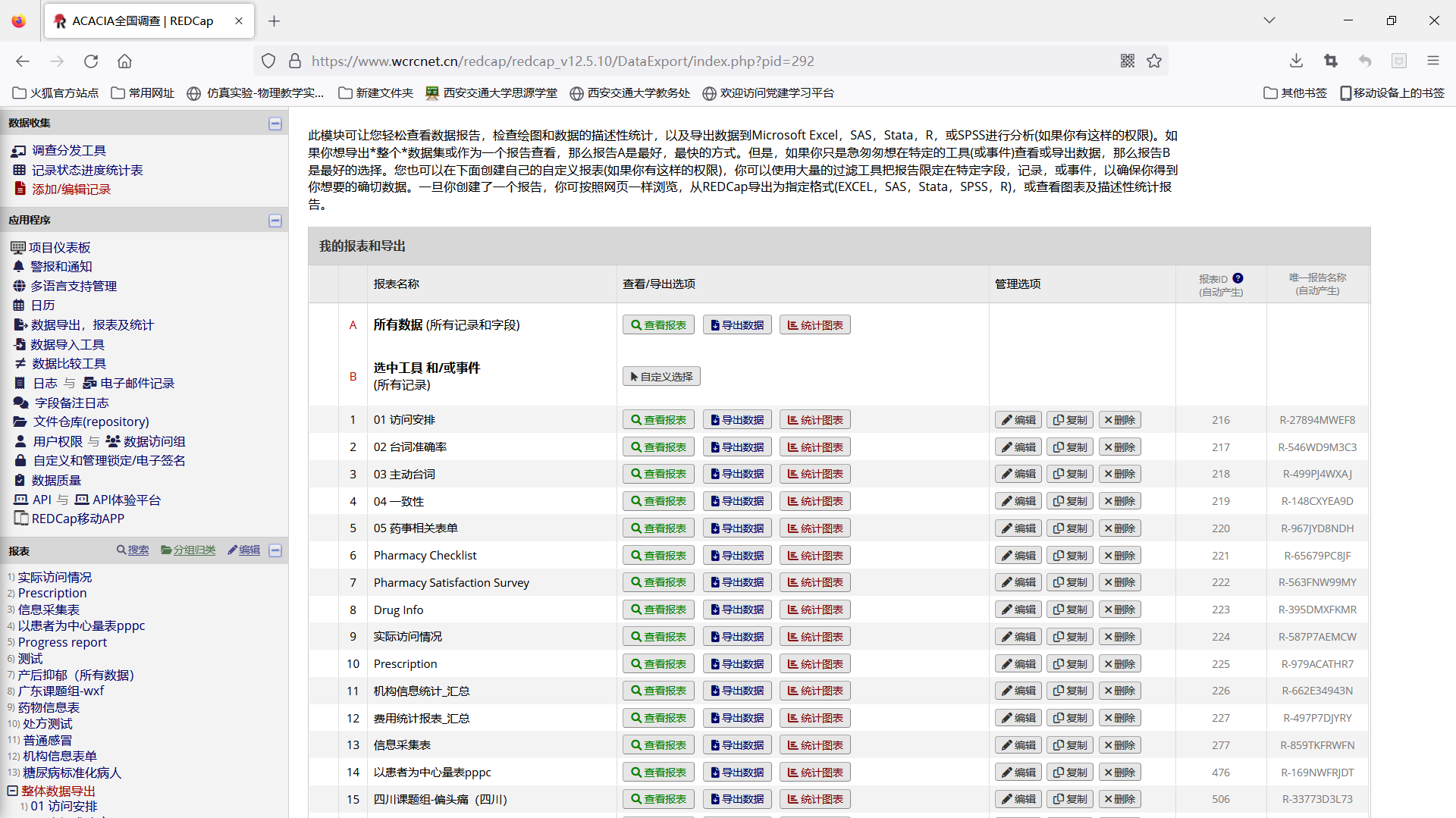
图1-2 项目主界面

核心系统的项目主界面为四个部分：项目页面和项目设计、数据收集、应用程序、帮助及咨询（图1-2）。

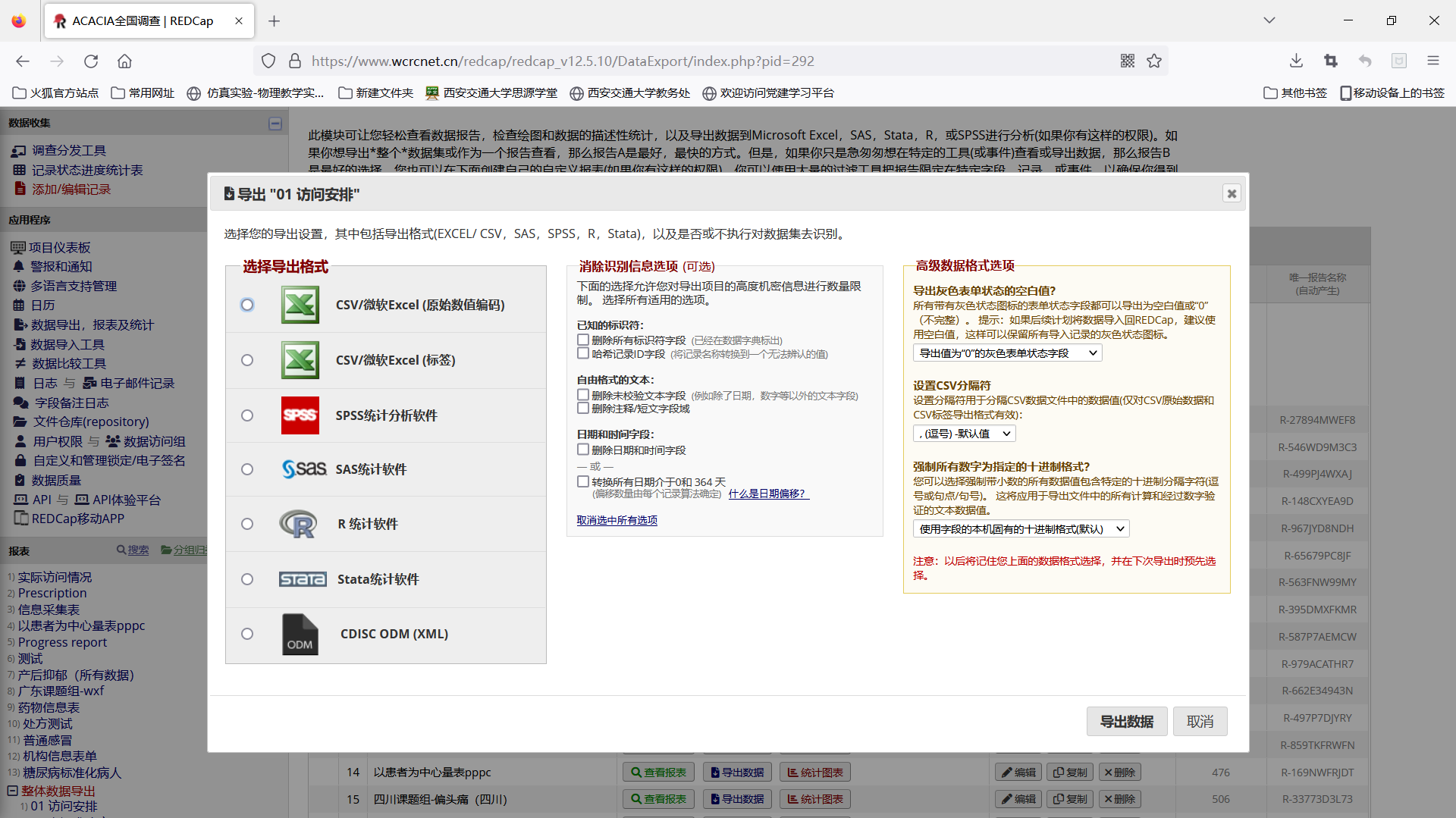
**2.操作步骤**

**2.1进行数据导出**

点击应用程序部分—“数据导出、报表及统计”（图2-1）， 分别对01访问安排、02台词准确率、04一致性进行数据导出，在“数据导出、报表及统计”模块中，可实现所有数据的查看和数据的导出。在“记录状态进度统计表”中只能看到单条记录的具体数据，而该模块中的“查看报表”功能，可将所有记录整合成为一个表单，快捷查看所有录入的记录。

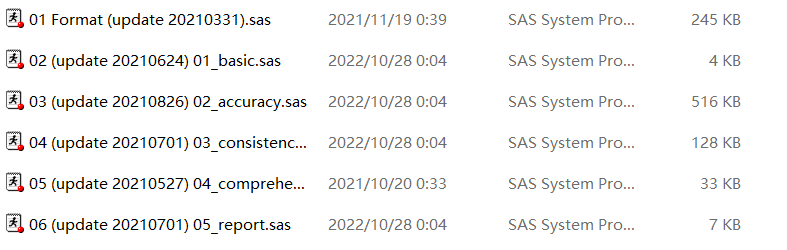
图 2-1 “数据导出、报表及统计”界面

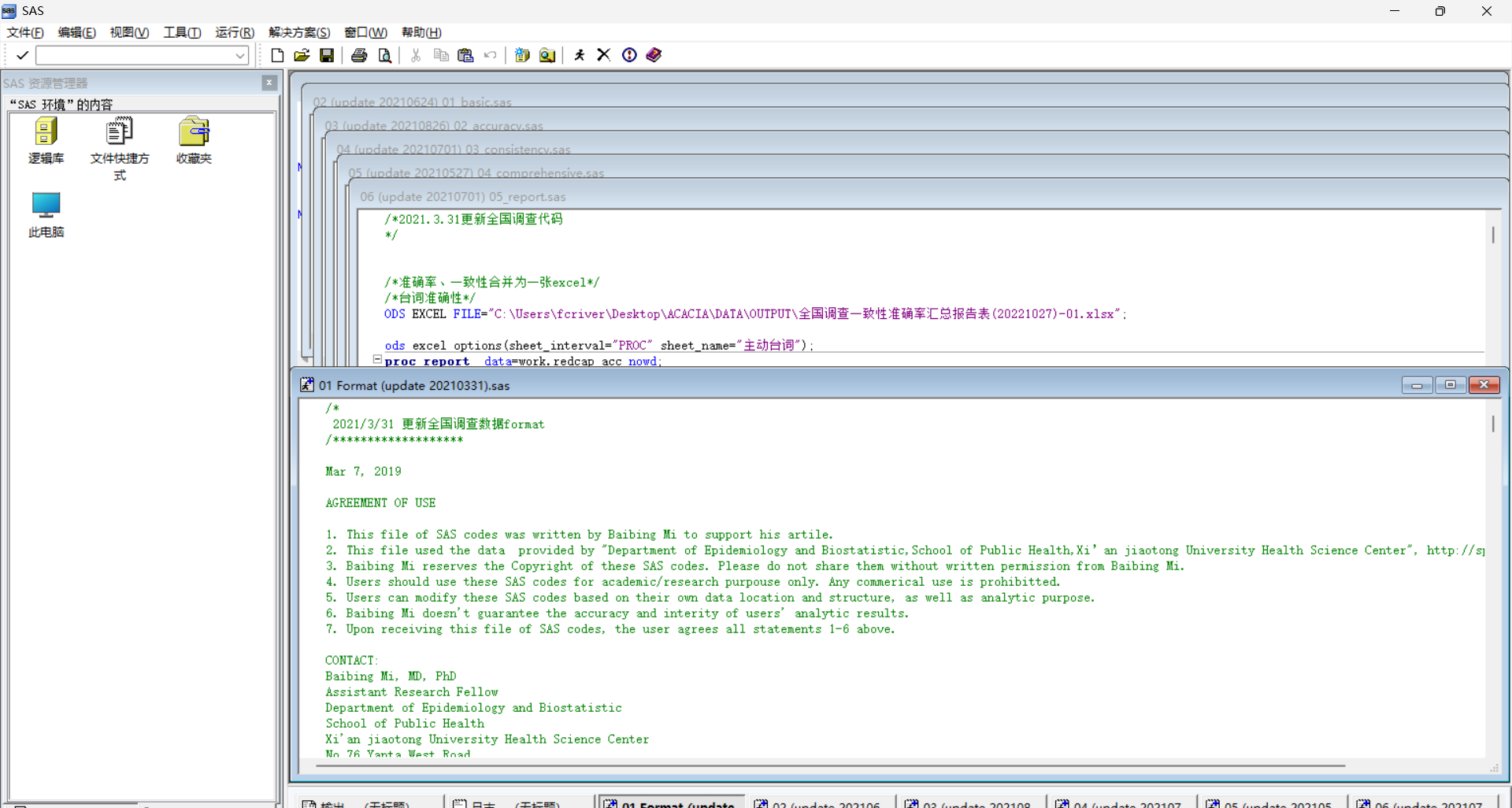
REDCap系统将录入的研究数据储存于网络，有权限的研究人员可利用“数据导出”功能，将录入数据导出至本地。系统支持导出为各种主流统计软件的格式，包括Excel、SPSS、SAS、R软件、Stata软件。根据后续分析计划安排，可选择相应的导出格式，并进行其他导出设置。将数据导出为所需的特定格式，可进一步在其他统计软件中对数据进行检查与整理，保证数据质量（图 2-2）。

图 2-2 导出数据将录入的数据导出至本地

**2.3使用SAS9.4进行数据质控**

依次打开编写的代码（图2-3）01 Format (update 20210331) 、02 (update 20210624) 01\_basic、 03 (update 20210826) 02\_accuracy 、04 (update 20210701) 03\_consistency 、05 (update 20210527) 04\_comprehensive 、06 (update 20210701) 05\_report 打开（图2-4）。

图2-3“依次打开code1-6

图 2-4 打开代码后SAS界面

01 Format (update 20210331)的功能为指定数据存储路径，并为下一步的数据指定统一分析格式；（图2-5）

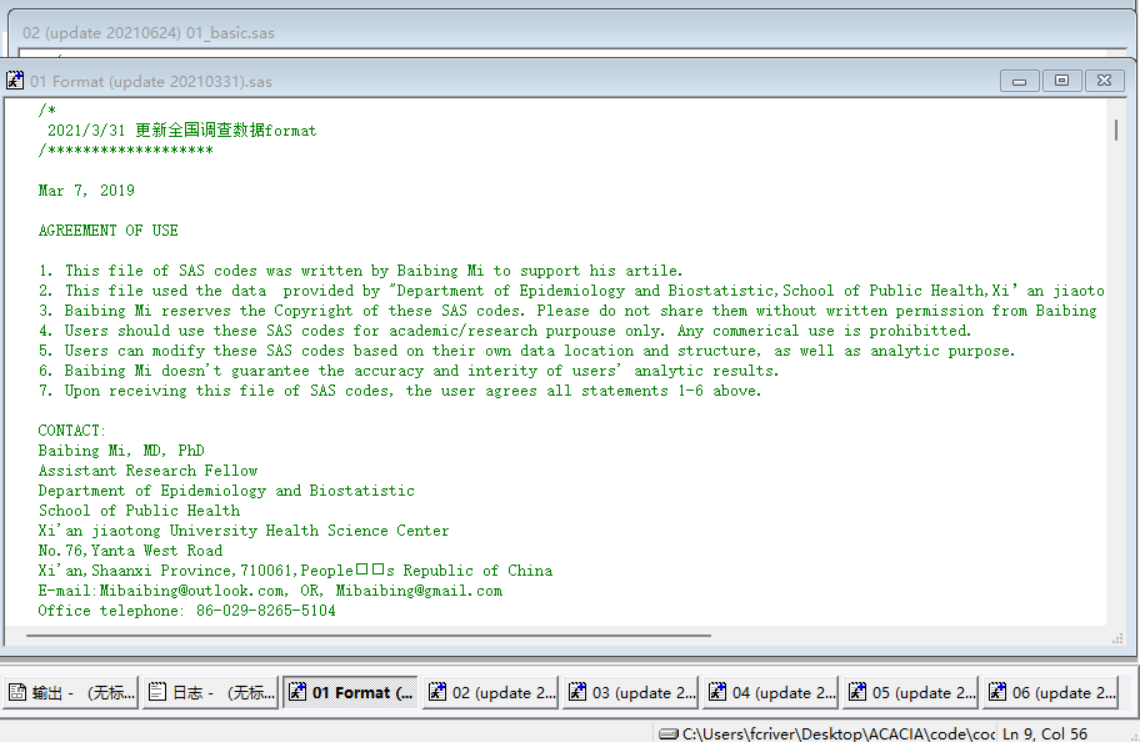


图 2-5 01 Format

将导出的数据文件ACACIA-01\_DATA\_NOHDRS\_2022-11-04\_2310.csv导入代码02、该步02 (update 20210624) 01\_basic的功能为导入标签数据与人员访问安排并对数据按编号进行排序；（图2-6）

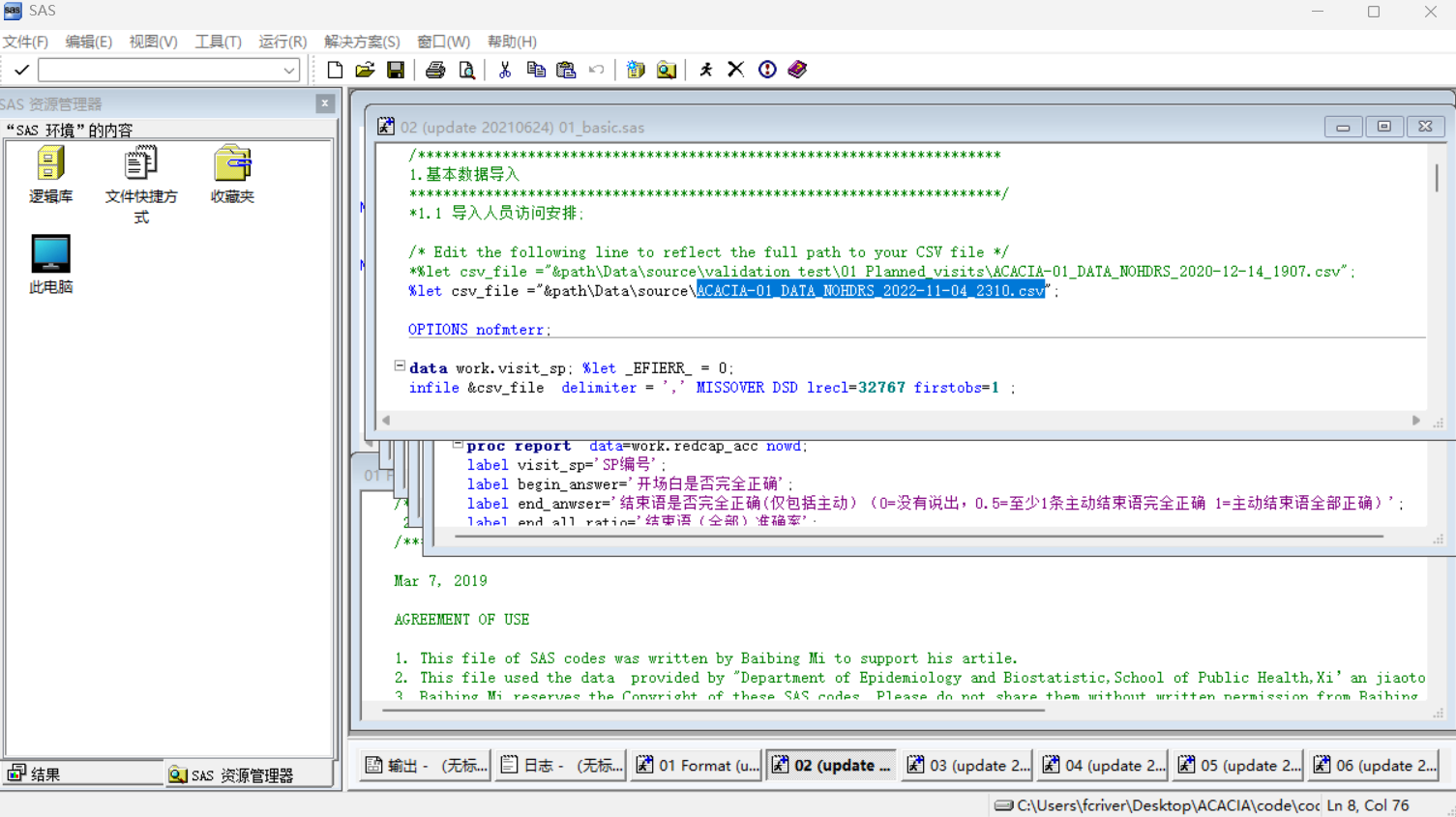


图2-6将原始数据导入代码02

将导出的数据文件ACACIA-02\_DATA\_NOHDRS\_2022-11-04\_2310.csv导入代码03、该步03 (update 20210826) 02\_accuracy的功能为导入计算台词准确率的数据，赋予标签加以排序，并将各组疾病分类分别构建新变量,计算台词准确率，最后纵向合并数据集；（图2-7）；

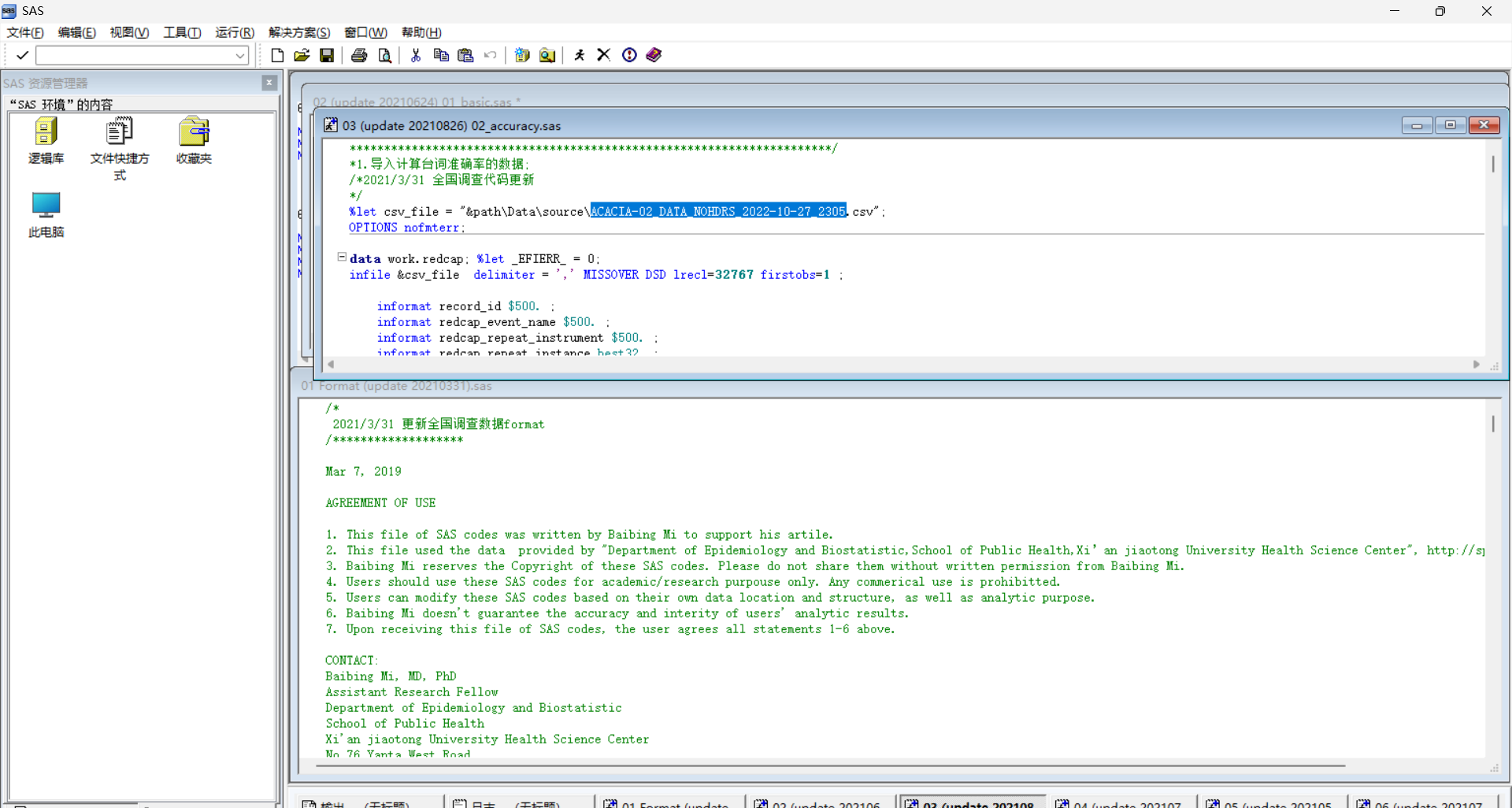


图2-7将原始数据导入代码03

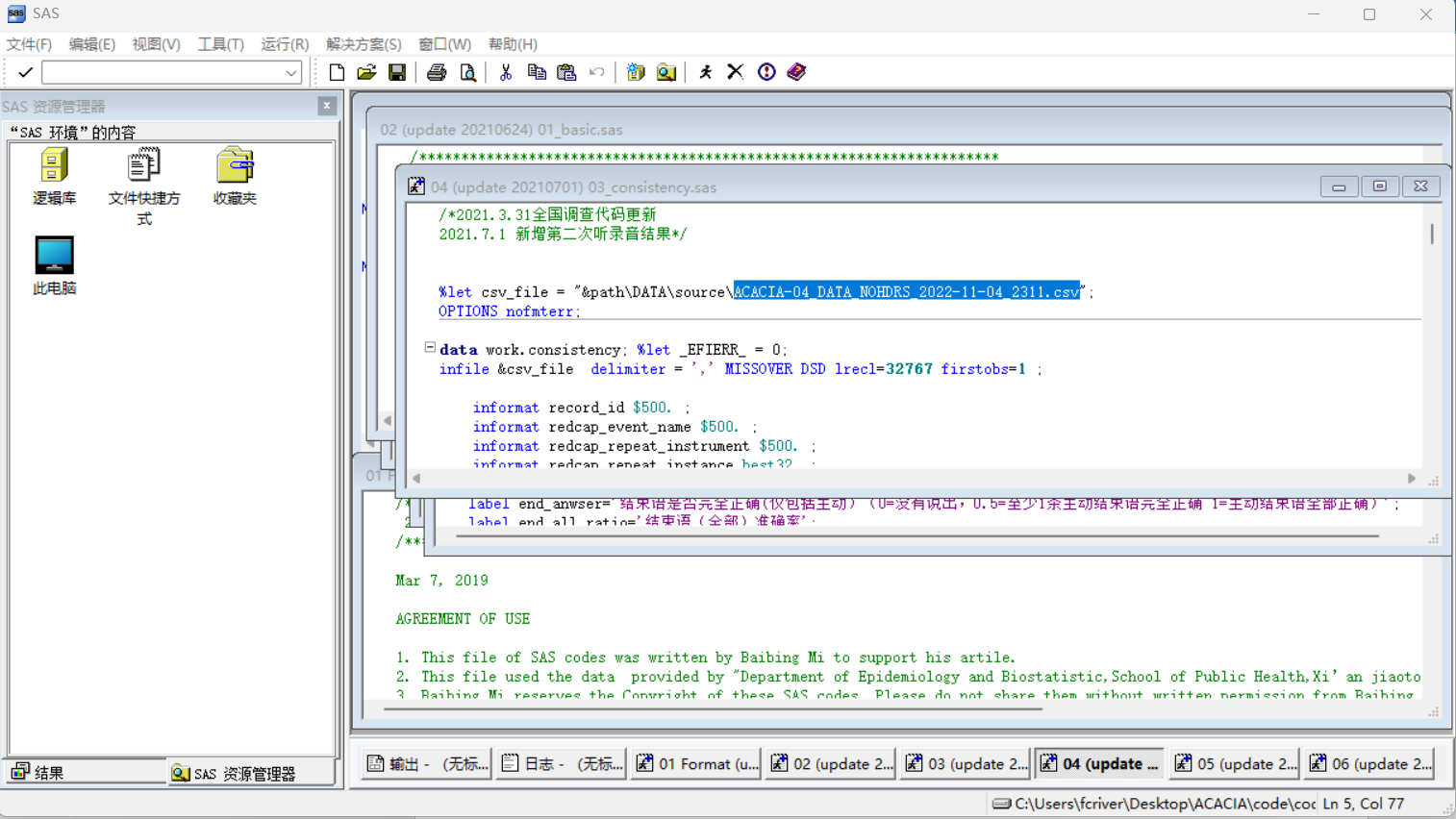
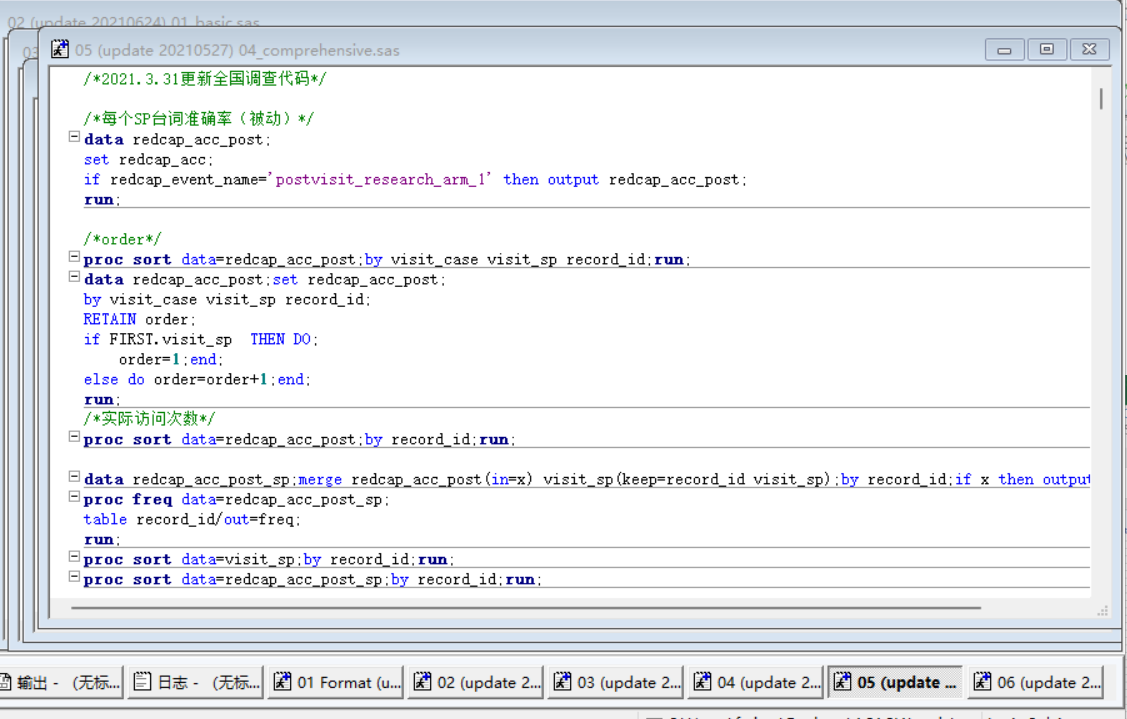
将导出的数据文件ACACIA-04\_DATA\_NOHDRS\_2022-11-04\_2311.csv导入代码04、该步04 (update 20210701) 03\_consistency的功能为导入一致性评价数据并赋予标签，根据病种保留特定数据展示数据录入情况，通过转置对一致与不一致进行对比，根据两次录音结果对数据进行调整，并对录音质量进行评价，得出一致性描述性统计结果；（图2-8）

图2-8将原始数据导入代码04

05 (update 20210527) 04\_comprehensive的功能为计算SP台词准确率与实际访问次数，调整成输出格式，计算全部台词、被动台词的平均准确率，根据平均被动台词与全部台词准确率计算是否合格并标注省份, （图2-9）

图2-9 05 (update 20210527) 04\_comprehensive

06 (update 20210701) 05\_report的功能为将上述各数据赋予标签并汇总合并为Excel；（图2-10）

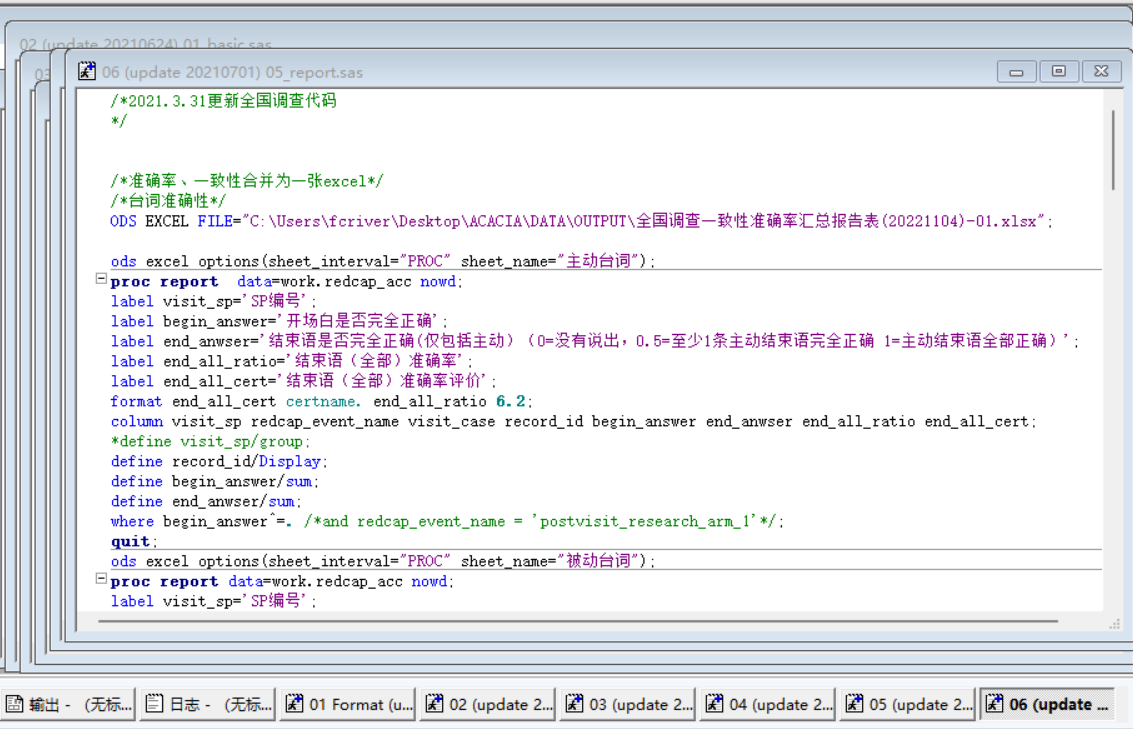


图2-10 06 (update 20210701) 05\_report

导入代码02、03、04，(图2-6、图2-7、图2-8)并依次运行代码01至06，结果如下。（图2-11、2-12）

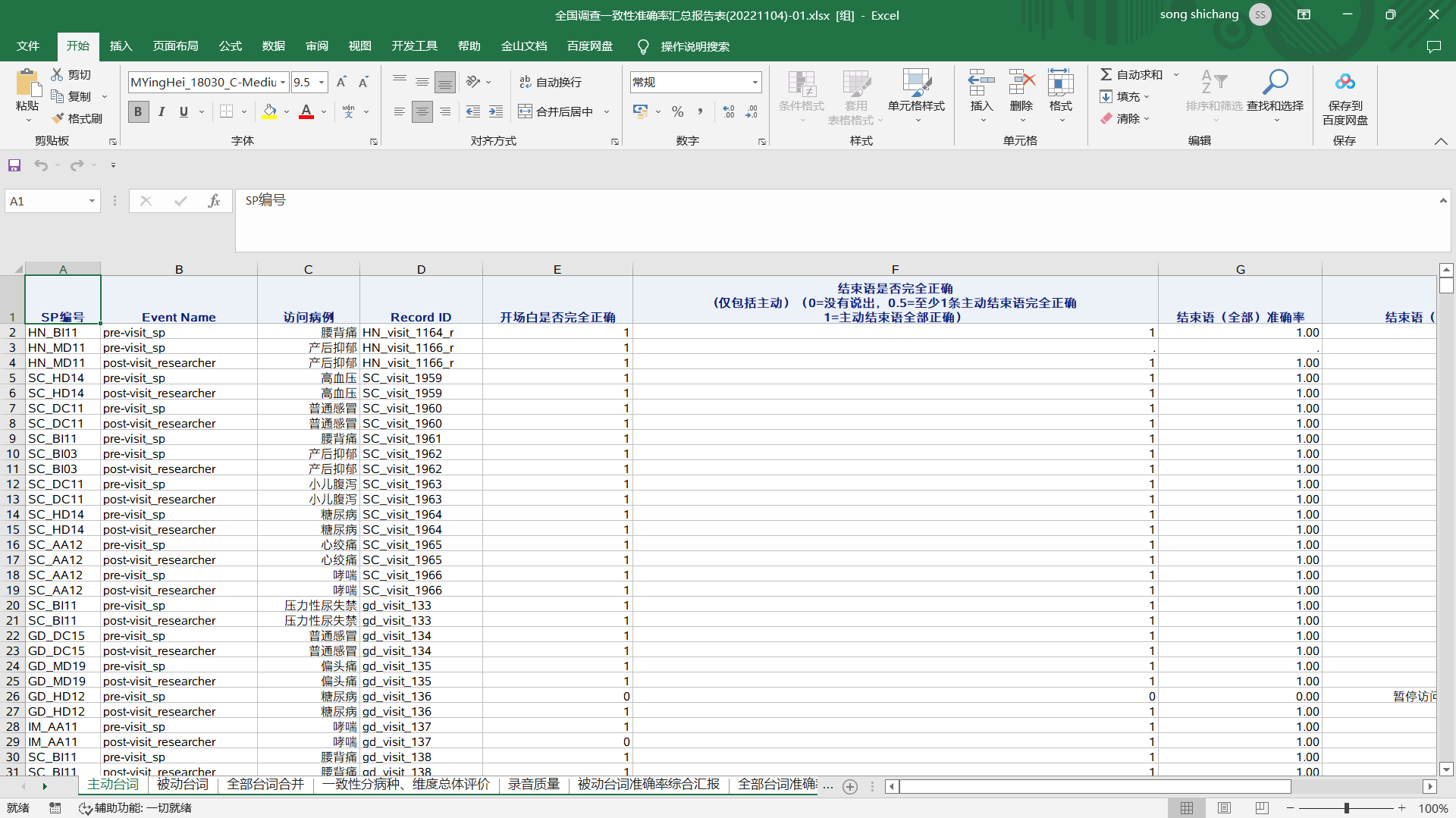


图 2-11 全国调查一致性准确率汇总报告表(20221104)-01

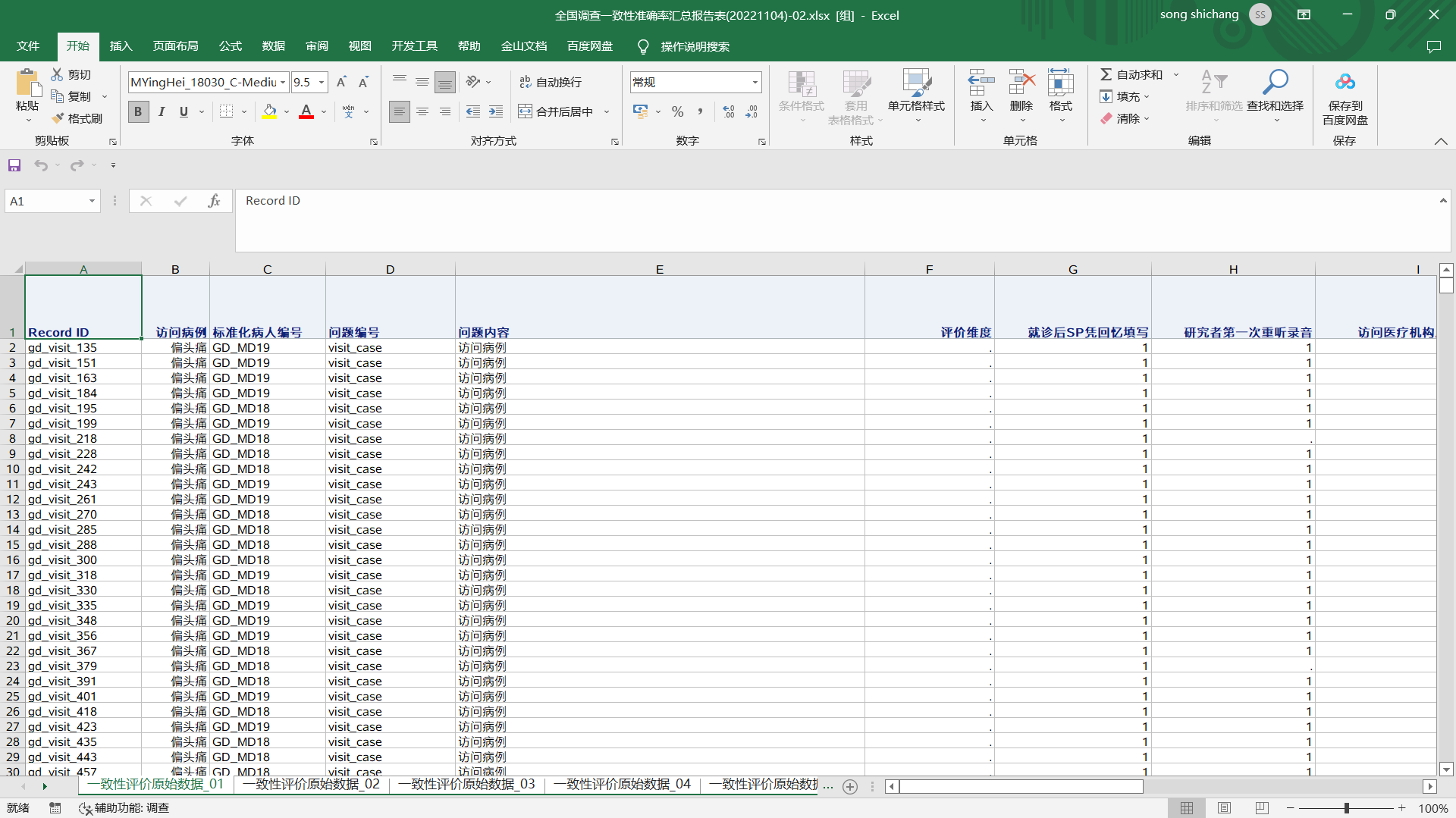


图 2-12 全国调查一致性准确率汇总报告表(20221104)-02