编码手册

**1 编码手册说明**

本册旨在规范编码流程，减少编码者之间的不一致性。

**2 编码流程**

编码流程主要分为编码阶段和校对阶段。

**2.1 编码阶段**

145篇PS2022文章和XXX篇大团队科学文章

对于那些使用大型数据集和二手数据的研究、动物研究、个案研究以及数据挖掘类研究，不查找或提取其人口学变量的公开数据，且编码时只摘取其简要信息。

阅读方法部分

查看文章是否公开数据，对于未提供原始数据的文章，联系作者获取数据

合并人口学变量数据

提取这些文章的人口学变量数据

筛选出可获取人口学变量原始数据的文章

编码每篇文章的完整指标

图1 编码阶段流程图

2.1.1 阅读方法部分

首先，我们阅读方法部分以判断研究的数据是否来自于大型数据集或是二手数据、或者是不是动物研究、个案研究或数据挖掘类研究。如果是，我们仅记录其简要信息（见文件“code\_example”第一个表格的第二条编码）。如果不是，我们将会根据本手册的第三部分摘录这篇文章的完整指标（见文件“code\_example”的第一个表格和第二个表格的第一条编码）。

2.1.2提取人口学变量的原始数据

对于可获取原始数据（公开获取或通过给作者发邮件获取）的文章，我们提取每篇文章的人口学变量数据，对每个研究整理出一个只包含被试ID和人口学变量的数据文件，最后再把PS2022和大团队科学文章的数据分别合并成一个汇总数据集，以便后续的数据分析。

**2.2 编码校对**

（1）为了进一步增加编码内容的可靠性和客观性，在完成编码阶段后，我  
们将重新分配文章进行校对。

（2） 编码者分成若干组，每组有 2 名编码者对同样的文献进行校对，若编码内容出现分歧，回到原文献对比，并商定结果；若商定无法达成一致，由第三方最终裁定。

**3 编码详解[[1]](#footnote-1)**

* **文章编号**

每篇文章都有唯一的ID（8位数）。

* **题名**

文章的标题。

* **研究编号：**

说明：1.研究编号指摘录的编码信息来自文章的第几个研究；2.如果文章中有预研究，只有预研究以正式研究的形式书写才进行记录。该情况下，预研究编码为1，其他研究的编号递增。3.如果一篇文章中不同研究使用同一批被试，只需要记录该文章的文章编号、题名、研究编号和研究类型，其余编码内容无需填写，并且在remark部分备注：被试重复使用。

* **研究组别：**

说明：如果同一研究中被试被分成若干组（比如：青年组 vs 老年组），  
且分组依据是被试的人口学信息， 请分别在不同行记录被试的信息。

* **数据可获取：**
  + 0.不可获取； 1.可获取
* **研究类型：**
  + 1.实验/准实验； 2.问卷； 3.质性研究； 4.其他
* **样本类型：**
  + 1. 大学生（含硕博研究生）； 2.非大学生的学生； 3.婴幼儿； 4.学龄前儿童； 5.非学生的成人； 6.其他
  + 摘取文章中关于描述被试类型的句子或文字。

说明： 1.婴儿的年龄范围为 0~1 岁，幼儿的年龄范围为 1~3 岁，学龄前儿童的年龄范围为 3~6 岁。 2.若有研究同时涵盖各类群体或未明确说明，则记为其他。 3.若为追踪研究，被试在不同追踪时间上身份有所转变，样本的类型为研究所陈述的主旨人群为主，若无法判断则为被试最后一次参与研究时的所属类型，其余人口学信息也以最后一次为准。 4.在职硕博研究生需编码为“4.非学生的成人”。 5.摘取原文信息时请务必保证是完整的句子且句子能体现出被试的类型。

* **样本量：**
  + 编码：若报告，记录样本量，若未报告，标记为X。

说明：1.样本大小指的是研究的有效被试量。2.编码时若文章同时具有表格和文字描述，以文字描述为准。3.编码时若文章文字描述与数据集不一致，需仔细审查数据与文章内容，以剔除无效数据后进入主要分析的有效被试量为准。4.若有效被试是按照不同实验组或者不同群体进行描述，请将各组或各群体求和，然后记录。

* **性别：**
  + 0. 文章或补充材料中未报告，且原始数据不可获取； 1.文章或补充材料中报告，但原始数据不可获取； 2.原始数据可获取
  + 如果文章或补充材料中报告，但原始数据不可获取，则记录男性人数（占比） /女性人数（占比）/其它性别人数（占比）

说明：1.如果文章描述上同时拥有男性、女性和其它性别人数以及各性别占比，请记录人数。 2.若文章仅提供男性、女性和其它性别人数占比，请分别记录占比。

* **年龄：**
  + 0.文章或补充材料中未报告，且原始数据不可获取； 1.文章或补充材料中报告，但原始数据不可获取； 2.原始数据可获取
  + 如果文章或补充材料中报告，但原始数据不可获取，请记录年龄的详细信息

说明：1.若文章中的年龄信息既未报告M±SD，也未报告年龄范围， 请记录文章中具体的描述内容。例如：13及以下：28.9%；14~16：46.2%；17 及以上：24.9%。 2.如果文章同时报告年龄的多种指标，尽可能都进行记录。

* **社会经济地位：**
  + 0.文章或补充材料中未报告，且原始数据不可获取； 1.文章或补充材料中报告，但原始数据不可获取； 2.原始数据可获取
  + 如果文章或补充材料中报告，但原始数据不可获取，请记录其报告的信息类别

说明：社会经济地位的信息类别分为： 1. 收入或者其他财产； 2.主观社会阶层。记录时只需要记录相对应的数字即可。若被试同时涵盖以上两个类别，请记录为：12。

* **受教育水平：**
  + 0.文章或补充材料中未报告，且原始数据不可获取； 1.文章或补充材料中报告，但原始数据不可获取； 2.原始数据可获取
  + 如果文章或补充材料中报告，但原始数据不可获取，请记录受教育水平的详细信息

说明：只有当样本类型为 4 或 5 时才填写该项编码。

* **种族（race/ethnicity）：**
  + 0. 文章或补充材料中未报告，且原始数据不可获取； 1.文章或补充材料中报告，但原始数据不可获取； 2.原始数据可获取
  + 如果文章或补充材料中报告，但原始数据不可获取，请记录其信息，例如：white; black。
* **职业：**
  + 0. 文章或补充材料中未报告，且原始数据不可获取； 1.文章或补充材料中报告，但原始数据不可获取； 2.原始数据可获取

说明：只有当样本类型为 4 或 5 时才填写该项编码。

* **宗教信仰：**
  + 0. 文章或补充材料中未报告，且原始数据不可获取； 1.文章或补充材料中报告，但原始数据不可获取； 2.原始数据可获取
  + 如果文章或补充材料中报告，但原始数据不可获取，请记录其报告的信息类别。

说明：宗教信仰的信息类别分为： 1.佛教； 2.基督教； 3.伊斯兰教； 4.其他宗教信仰。记录时只需要记录相对应的数字即可。

* **国家或地区：**
  + 0. 文章或补充材料中未报告，且原始数据不可获取； 1.文章或补充材料中报告，但原始数据不可获取； 2.原始数据可获取
  + 如果文章或补充材料中报告，但原始数据不可获取，请记录其详细信息。

说明： 1. 国家或地区的编码记录为ISO 3166-1二位字母代码，比如：中国记为CN； 2.如果一个研究包含多个国家或地区，标注了具体国家/地区的每个国家或地区单独记录，3.如果一个研究没标注具体的国家或地区，只标注了更大范围区域的，直接记录这一区域，例如，“欧洲”。

* **抽样方法：**
  + 0.未报告； 1.方便抽样； 2.随机抽样； 3.其他
* **被试招募方式：**
  + 0.未报告； 1.报告
  + 如果报告，记录不同的被试招募方式。其中，在线数据收集记录在线  
    数据收集的平台。如果平台名无法获取，记为“crowdsourcing”。线  
    下数据收集记为“offline”。不同的被试招募方式使用英文状态下的  
    逗号隔开。
* **摘要：**
  + 是否报告被试信息： 0.未报告； 1.报告
  + 如果被试为大学生，是否在摘要中提及： 0.未提及； 1.提及。
* **generalization statement：**
  + 目标人群： 1.陈述的特定人群；2.陈述的一般人群； 3.推断的一般人群。
  + 目标人群编码依据：记录编码者做出目标人群编码的依据，即摘取原文。
  + 限制可推广性：0.无相关描述；1.有相关描述
  + 限制可推广性编码依据：记录编码者做出限制可推广性编码的依据，即摘取原文。

说明： 1.对于目标人群和对进一步推广局限的描述的编码主要阅读文章的讨论和结论部分，但文章标题、摘要等部分有时也能做明显判断，编码者可综合多部分信息快速进行判断。2.对于目标人群的编码，下面提供一个决策树供大家理解。

是

陈述的特定人群

是否是特定人群

否

推断的一般人群

否

是否直接使用“人类”等一般性词语进行描述

是

陈述的一般人群

图2 目标人群决策树

* **备注：**

说明： 1.大型数据集、二手数据、个案研究、动物研究以及挖掘数据类  
研究请分别备注为： database， secondary data， case， animal， mining data； 2.多条备注请在每条备注前添加序号。 3.除了必须备注之外，编码人员可以备注自己认为编码时想要记录的事项。

1. 若在同一编码中，需要记录多条信息，信息间使用英文状态下的分号隔开。 [↑](#footnote-ref-1)