

1981-1983 五城市调查最终调查数据产生始末

中国社会学重建之后于 1981-1983 年进行的五城市调查，是中国社会学抽样调查的起点。它始于 1981 年在天津尖山街进行的婚姻家庭和生育问题调查。那时中国社会学会接待了受美中学术交流学会资助来华访问的美国人类学家，美国纽约市立大学人类学教授帕斯特内克（Burton Pasternak）。应他的要求费孝通先生安排他到天津进行实地调查。调查的题目是婚姻家庭和生育问题，调查的对象为调查所在地全部 55 岁以上的已婚妇女。调查采取的是入户访谈式的问卷调查。中国社会科学院社会学研究所的潘乃谷和天津社会科学院社会学研究所的潘允康组织实施了此项调查。

当时，费孝通先生认为，中国社会学正处在起步阶段，百废待兴，要认真学习西方国家的经验。他认为西方的经验不仅是理论方法，同时也包括如何做调查。他在天津调查结束后对那时所里正在北京宣武区椿树街道的东河沿居委会进行社会调查基地建设的人（我和所里的同事薛寅、张仙桥和王荣芬等）说不入虎穴焉得虎子，我们也要在东河沿做一个同样的调查，把他们的经验学到手。于是有了 1982 年的东河沿婚姻家庭和生育调查。在接受任务以后我们参照天津调查的问卷设计了东河沿调查的问卷。为了有可比性，我们保留了天津调查绝大部分项目，但是也根据研究的需要，增加了不少新的项目，所以东河沿调查的内容比尖山街问卷调查篇幅大不少。调查对象也从 55 岁以上的已婚妇女，改为全体已婚妇女。和天津调查一样，东河沿调查也采用入户访谈的方式。

整个东河沿调查都是在费孝通先生的指导下进行的。参加调查的人几次到他家中开会，那时他的家还在中央民族学院内，是一个两居室的单元房。地方狭窄，只有两个卧室。他就坐在卧室中的一张八仙桌旁，听取我们的汇报并对调查工作做出指示。他提出调查人员不仅要调查，而且要利用调查得到的资料写文章。根据他的要求，我用东河沿的资料，写了一篇论婚姻的当户对的论文，他亲自审阅，批示，认为论文很好，有理有据，指出社会学者论文就是应该从调查资料出发来写文章。在此期间，费孝通先生还在家中安排了对《中国青年生育意愿调查》的讨论。希望我们能利用东河沿调查的资料对它做一些补充。实际上因为那时《中国青年生育意愿调查》这一项目的实地调查已经完成，已经进入了分析写作的阶段，所以东河沿调查的资料实际上并没有起到什么补充作用，与会的人只是对分析和写作的框架提了一些意见，这些意见中张之毅先生的意见是非常有分量的，对最终报告的完成具有指导意义。

如果说东河沿调查对家庭研究有什么发现的话，那就是在东河沿调查中，我们第一次提出了用与被调查人共同生活这一指标作为“家庭”这一概念的操作定义。在东河沿调查最初进行的访谈中，我们发现在问及家庭人口的时候，如果不加任何限制，被调查人常常会将调查时健在的夫方所有的男性亲属及其配偶包括进来，家庭人口一般都会有十几口人，有时甚至会有几十口人。这固然反映了当时中国人心目中“家”这一概念，但显然这不是学术上的“家庭”的概念。因此，在遇到这样的情况时我都会进一步询问，在这些人中，哪些人现在和你一起过日子？这时被调查人的家庭人口便会迅速减少，因为那些不在一起生活的人此时都会被排除出去，这就使家庭人口回归到“家”的理论定义限定的范围。在调查中我们还发现，对“共同生活”这一概念我们不必做出硬性规定，如是否住在一起、是否在一起吃饭等，只要由被调查人主观认定就可以了。换言之，东河沿调查既没有使用家庭社会学经典的关于“屋顶”（one roof one family）的操作定义，也没有使用费孝通先生“一起吃饭”（同灶）这样的标准，而是采用了主观认定的“共同生活”这一标准作为家庭的操作定义。之后其他调查点进行的调查，以及我以后所做的家庭调查都采用它作为家庭的操作定义。东河沿调查使

我认识到在家庭研究中怎样在操作层面定义一个家，是一个十分重要的问题，

1982年夏，中国社会学会第一届年会在武汉的华中工学院，即今天的华中理工大学举行。我应费孝通先生之命，在会上做了一个题为《论基地》的报告。关于“调查基地”的思想是费孝通先生一贯提倡和坚持的社会学者做学问之道。他认为社会学者不能脱离社会，必须时刻关注社会的变化，而社会调查基地正是社会学者观察和了解社会的窗口。正是在费孝通先生的倡导和主持下，建立了社会学恢复之后的第一个调查基地——北京宣武区椿树街道东河沿居委会调查基地，即东河沿调查基地。而东河沿婚姻家庭和生育调查则是在那个基地完成的第一个调查项目。所以我在年会上的报告除了阐述自己对费孝通先生关于社会调查基地思想的理解以及在东河沿进行社会调查基地建设的情况外，也以相当的篇幅对东河沿的婚姻家庭调查做了介绍。报告引起了与会者热烈的反响。于是在年会期间，费孝通先生趁热打铁，在自己下榻的宾馆的房间举行了一次由与会各单位有关人员自愿参加的吹风会，社会学所参加这次会议的有潘乃谷和我，其他与会者我记得有天津社会科学院的潘云康、江苏公安专科学校校长姬仰曾，四川社科院社会学所所长詹一之、上海社科院社会学所薛素珍、北京经济学院的王胜泉和上海高校的与会者等十几个人。他向与会者介绍了东河沿调查的情况，提出在其他城市进行同样调查的建议。他明确指出，参加不参加，由各单位自己决定，姜太公钓鱼愿者上钩，参加的单位自筹调查调查经费和差旅费，中国社科院社会学研究所可出具证明文件，承担会议和数据处理费用，以及组织协调与处理数据等技术工作。会议气氛很热烈，与会者都表现出强烈的参与愿望。姬仰曾和詹一之本身就是单位的负责人，所以当即就拍板参加，另一些则表示自己愿意参加，但须回去请示领导后再定。会后，各单位很快就有了正式的回音。最终确定参加的单位有：中国社会科学院社会学研究所、北京经济学院、天津社会科学院社会学研究所、江苏省公安专科学校、四川省社会科学院社会学研究所、成都市社会科学研究所以、上海社会科学院社会学研究所、上海复旦大学和上海复旦大学分校等（后来的上海大学）。当时叫做联合调查组。

联合调查组成立后的第一项工作就是设计用于该项调查的问卷。这一项工作是在北京的东方饭店举行的。与会者以东河沿调查的问卷为蓝本，逐项讨论斟酌，形成了五城市联合调查组使用的统一问卷。调查并没有规定统一的名称，由各参加单位根据实际情况自行确定。各个单位的调查点也是由各个单位根据情况自行选定，而非由事先规定的程序统一抽取的。最后选定并实施的调查地点是：北京宣武区椿树街道东河沿居委会、北京朝阳区团结湖街道团结湖居委会、天津河西区尖山街道红星第二居委会、上海市徐汇区新乐街道张家弄居委会、上海杨浦区宁国街道双阳居委会、上海虹口区长春街道、南京夫子庙街道四福巷居委会和成都如是庵街居民段。在调查对象的选择方面，除上海虹口区长春街道采用的是街道内的简单随机抽样外，其余的都采用了整群抽样，即调查对象是调查地（抽取的户中）全体已婚妇女。除已经先期完成的天津尖山街道红星居委会和北京东河沿之外，其余调查都是在1983年初实施的。所有调查都是采用调查员入户访谈这样的方法来填答问卷。

1983年，在全国社会科学规划会上，这一调查被确定为“六五”全国哲学与社会科学的重点项目，项目冠之以“中国城市家庭研究”这一名称，但一般习惯上我们仍叫它五城市调查。项目聘请了雷洁琼先生担任学术指导。费孝通先生则改任另一重点项目，“小城镇研究”的学术指导。最后确定的项目组成员有：中国社会科学院社会学研究所刘英（项目负责人）、沈崇麟；北京经济学院经济系刘金云；天津社会科学院社会学研究所潘云康、（项目负责人）、张雅芳、上海社会科学院社会学研究所薛素珍（项目负责人）、杨善华；上海大学文学院社会学系刘炳福、仇立平；上海复旦大学哲学系范伟达；江苏省公安专科学校夏文信、宋践；四川省社会科学院社会学研究所李东山、成都市社会科学研究所以吴本雪。那时，早前联合调查组有些人，比如沈关宝，记得在东方饭店讨论问卷的会上，我和他为了家庭的操作定义曾争得脸红耳赤，但等项目组成立时他已经离开，随费孝通先生去了小城镇项目。而复

旦大学分校也成为了上海大学的一部分。项目实施过程中，作为学术指导，雷洁琼先生给予了我们极大的支持，她将自己的经验学识倾囊相授。她不仅参加了项目组在北京的全部会议，而且还参加了项目组的连云港会议和厦门会议。每一次会议都会亲自参与项目组的讨论并提出自己的意见和建议。可以说没有雷洁琼先生的指导就不会有项目的顺利完成。她的“要用资料说话”、“论文写作要规范”和“社会学要面向中国的实际”（杨善华、王思斌《雷洁琼学术生涯二三事》）的学术研究导向对我今后的研究有很大的影响。五城市调查之后，在我参加和主持各个家庭调查中项目，都请她担任学术指导，每一次都得到了她的指导和支持。1996年中日合作进行的城乡家庭调查，90高龄的雷老还亲赴太仓，出席太仓调查的调查员动员会。不仅如此，雷老也给了东河沿调查基地以无形的支持。在费老因为大家知道的原因离开社会学所之后，东河沿调查基地处于无人问津自生自灭的境地，我出于一种责任和对东河沿的感情，还在那里苦苦支撑。除了平时常去那里看看之外，逢年过节都会自掏腰包和科研处的同事陈婴婴、赵克斌等一起给居委会送去一些可乐、挂历之类的东西。在调查开始或结束时，常常会请居委会的主任们聚餐，联络感情。这时我都会请雷老到场，在这样的场合，雷老与居委会的主任们谈天说地，聊家常，从不说官话套话。居委会主任们感到十分亲切，在他们眼里雷老不只是一位大领导，也是一位可亲可敬的老大姐。一次，在王府大街八面槽翠花楼的聚餐会上，居委会主任们对雷老说，请她以后还要来，雷老风趣地说“当然要来，有说有笑，还有吃”。雷老的话音一落，响起了一片笑声。而正是在这笑声中，东河沿基地得到了雷老无形的支持。使东河沿调查基地又延续了十余年。在此期间，在那里进行了七城市家庭调查的东河沿调查、接待了“亚太地区家庭未来讨论会”和“中日家庭比较研究”的与会代表。日本家庭研究者访问了东河沿的居民家庭，发现了一种在日本从来没有看到的家庭形式——隔代家庭。后来担任过英国牛津大学中国研究中心主任的彭轲教授的博士论文中的实地调查，有相当大一部分就是在那里做的。而这一切之所以得以实现，都离不开雷老的无形支持。

在这里，我也要向项目组历次会议的组织和资助者表示自己的谢意和敬意。他们是中国社会科学院社会学研究所的潘博儒、江苏省公安专科学校校长姬仰曾和《家庭》杂志社社长叶先生（时间久远，已经想不起名字）。

五城市调查的数据全部用计算机处理，这也是费孝通先生倡导的。东河沿调查之初，费孝通先生访美归来，召集调查组开会，谈了自己访美的观感，特别谈到在美国看到用计算机处理调查数据的感受，他用带着浓重的吴江口音的普通话说道，只看见数字（在屏幕上）蹉蹉（飞快）地出来，这些数字不就是文章吗？有了它们还愁写不出文章？他又一次对我们说道，不入虎穴焉得虎子，叫调查组用计算机来处理东河沿的数据。当时所里既没有计算机，也没有计算机人才。于是他请了自己的弟弟，华中工学院自动化系仿真实验室的副教授费奇来帮忙，费奇又找了自己的助手刘国明。而我这个鸭子也就上架成了计算机“人才”。在费刘两位先生的指导下，我开始了计算机数据处理的前期工作，学习 BASIC 语言，编制编码手册，编码和过录（将问卷上的编码登录到一张张专门用于数据录入的专用表格上）。五城市调查虽然是后编码的，但是每一个调查变量的分类都是事先已经确定了的，所以编码手册的编制只需给每一个类别定值，确定所有的变量值在过录表，也就是将来在数据文件中的位置和字长（即占用的格数）即可。因为字长需事先规定，所以对于数字型变量，视其占据格数的不同，以 8 和 8 的组合作为最大的码值。而将 9 和 9 的组合定义为缺失值。在此期间，我们所迎来了几位第一届南开社会学专业的毕业生，其中一位叫谢文，被分到了家庭研究室，他很快就投入到项目的工作中。他提出在数据录入之后，再进行一次逻辑检验（跨变量的检查数据存在的错误）。于是我们又一起编制了一份逻辑检验的清单，递交给华中工学院费奇方面，请他们据此编制逻辑检验的程序。一切准备就绪，我和谢文一起，背着问卷和已经过录好的过录表，赶赴武昌华中工学院，在那里的自动化系仿真实验室的机房，用他们用扩展

BASIC 语言编制的录入程序和逻辑检验程序，完成了数据录入、逻辑检验和改错工作，获得了撰写调查报告所需的统计数据。东河沿调查是第一个进行计算机数据处理的调查，以后各地都照此办理，在华中工学院完成数据的录入、逻辑检验和改错工作，并产生将来写调查报告需要的的统计数据。当时没有像现在这样供数据处理的软件包，如 SPSS、SAS 和 STATA 等。数据处理的程序都是用费奇和刘国明先生，用扩展 BASIC 编写的程序进行的。

确切地讲，数据录入是分城市分卡进行的。即每一个调查点的数据，分成三卡（三套过录表）按卡一条一条（一行一行）记录地输入计算计算机，生成与三个卡（有关卡的问题可参阅《五城市联合调查组》的问卷）对应的数据文件，存储在一张 8 吋软盘上。所以五城市调查的原始数据的物理形式是 8 张 8 吋软盘，每一个软盘都有三个对应于三个卡的数据文件。数据文件使用的计算机码是 CROMCO Z80 系列，而非 IBM 系列的 ASCII 码。1983 年我用所里的科研经费结余，购置了一台由北京计算机研究所和计算机五厂生产的 BCMIII。从此社会学所和项目组有了自己的计算机。这台计算机也是 CROMCO Z80 系列的，使用的也是两个 8 吋软驱。年底，项目组在北京举行《中国城市家庭——五城市家庭调查报告和资料汇编》（山东人民出版社，1984 年，济南）一书的编辑和定稿会议时。我们就是用那台机器对五城市调查的数据进行分析，为报告提供所需的各种统计数字和表格。为了实现这一目的，我们首先对东河沿和尖山街的数据，按照后来联合调查组的数据格式，重新进行了编辑和调整，以使数据格式与联合调查组其他调查点一致，产生了格式统一的五城市 8 个调查点的数据。

1983 年 6 月，广州社会科学研究所的于幼军主动提出用五城市调查的问卷在广州进行一次调查，调查结束后，何颂扬先生携问卷和过录表来京，我们也是用这台计算机处理了广州的数据。这样五城市 8 个调查点的数据，就扩充成了六城市九个调查点的数据。由于这项工作完成在《中国城市家庭——五城市家庭调查报告和资料汇编》定稿之后，所以该调查的数据和报告并未收入该书。

1985 年出版的《中国城市家庭——五城市家庭调查报告和资料汇编》（五城市项目组，山东人民出版社，济南，1985 年）一书是项目组的共同成果。责任编辑房喜兰女士是一位资深编辑，独具慧眼看中了这一选题，承担起该书的编辑、出版和发行的全部工作。该书出版对国内外的社会学界发生了很大影响，对家庭社会学研究和抽样调查起到了很好的推动作用。1986 年我去荷兰求学，学位论文就是使用这本书刊登的资料。我记得我第一次去莱顿大学汉学院去见学校为我推荐的指导教师彭軻（Frank Pieke）的时候，他从办公桌的抽屉里拿出的了一本书，对我说，你是研究中国家庭的，一定要看这本书，这本书就是《中国城市家庭——五城市家庭调查报告和资料汇编》。当时我也很惊奇，没有想到山东人民出版社也将这本书发行到了国外。而实际上我的学位论文使用的最重要的数据都来自这本书。根据分析得到的被调查人分年龄组的家庭结构分布，我论证了中国人一生生活的家庭结构是随年龄变化而变化的，与西方不同，中国的核心家庭，并非婚姻的产物，而是夫方父系家庭这颗大树生长过程中生长出来的，主干家庭，而非联合家庭，也非核心家庭才是中国传统家庭的基本结构。在各种家庭结构类型构成中，由于传统的中国，大多数家庭都是多子女家庭，因此核心家庭所占比例始终都是最高的。并不存在现代化理论所谓的核心化趋势。学位论文得到他的高度赞扬，为此他专门给校方致函，对论文做了很好的评价。我在 ISS 的指导教师，将论文寄给了荷兰家庭研究会会长和跨大学人口研究所（NIDI）所长库耶（KOOYE）教授。他也对论文做了较高的评价，认为我的观点很有新意。他建议在此基础上作进一步的研究。因为分组数据，毕竟还是截面数据，如果能用纵向的数据来证明这一结论，便会有更大的说服力。正是根据他的建议，我在以后自己主持的调查中，如七城市家庭调查和国情调查的问卷中都设计包含了一张家庭生命过程表，用来收集被调查人从出生到调查进行时生活的家庭结构类型信息。为了给我今后的研究提供更多的帮助，他邀请我去他乡下的家中详细讨论一下这个问题，他说他家里有一个小型的家庭研究书籍和刊物的小型图书馆，可为这个研究提供

很多资料。由于在电话中没有听清约会的时间，结果还害得老先生和他的夫人在乡下的火车站白等了好几个小时。那时我已经在准备回国的各项事宜，在离开荷兰之前再也没有机会访问库耶先生的家。后来我的同事折晓叶去荷兰访问，专门去荷兰的乡下他家中，代我探望库耶先生。在完成 ISS 的学业后，我应邀到 NIDI 做为期三个月的访问，利用这三个月的时间，我对这本书料汇编部分的统计资料进行了二次分析，完成了一篇中国家庭研究的综述性报告。NIDI 的同事对这个研究表现出极大的兴趣，希望今后能一起用原始数据对中国的婚姻家庭进行进一步的研究。我现在还记得我的同事巴特雷马当时问我的问题“Where is the raw data?”我现在做的工作就在回答他 34 年前他提出的问题。杨，我希望现在我做的工作，能实现你三十年前想分析的那套资料愿望。如果有机会，我们希望能有机会像当年在 NIDI 那样，坐在计算机前一起来分析它。

1990 年，我受社会学研究所委托，组织了联合国教科文组织委托的《亚太地区家庭未来讨论会》。这个会对于我来讲，有两大收获。首先，我在会上以《中国城市家庭——五城市家庭调查报告和资料汇编》和人口普查数据为依据，探讨人口要素对家庭结构的影响。我先用五城市分年龄组的家庭结构类型的分布，论证了传统中国家庭结构主干家庭为干，核心家庭为枝的结构形态。而支撑这种形态的人口要素是出生与死亡达到某种相对的平衡，即替代水平不低于 2.1。但是人口普查的数据表明，一胎化后中国人口替代水平已经低于 2.1，而且呈进一步下降的趋势。这意味着支持主干家庭为干，核心家庭为枝这种结构形态的人口要素已经不复存在，因而家庭结构形态必将随之发生变化，而这一变化，必将对我国社会其他方面产生巨大影响。当时我的报告的评论员是当时日本家庭社会学会的会长森冈清美（森冈先生是我的同事张萍推荐的）。森冈先生不仅对我提出的观点很感兴趣，而且也对产生我使用的数据的调查——五城市调查感兴趣。会后森冈先生对我说，以前对中国的家庭研究不太了解，现在才知道，原来中国方面已经做过这样大规模的家庭调查。他说这是中日双方今后开展合作研究的基础。他表示，回去以后很快就会组织一个老中青三代的研究团队，开展中日家庭的比较研究。森冈先生果然兑现了自己承诺，次年就率领一个由老中青三代组成的日本家庭社会学家组成的代表团，在北京和杭州两地举行了中日家庭比较研讨会。在杭州，森冈先生的学生，后来的日本家庭社会学会长，东京都立大学教授石原邦雄来到我的房间，由日本研究所的王伟先生做翻译，讨论了开展合作研究的问题，开展合作研究的口头合作协议，约定双方合作开展中日家庭的比较研究，双方共同讨论决定调查问卷和入户访谈提纲，由日方申请调查和研究经费，中方负责选择调查地，抽样和数据处理，共享调查数据。自此开始了我和日方团队延续 20 年的合作研究。在此期间进行了两次抽样调查、一次深入访谈和多次互访和讨论会。后来在石原先生的推动下，又吸收了韩国家庭研究者的加入，将中日比较研究，发展成中日韩的东亚家庭比较研究。今天活跃在社会学研究和教育第一线的老师，清华大学的王天夫、南京大学的吴愈晓、中央财经大学的范欢欢和中央民族大学的王纪芒等人的学位论文数据都来自这些调查。社会学所的同事陈婴婴和日本研究所的王伟都对这项持续 20 年的合作研究做出了很大贡献。

另外，我与所长陆学艺达成一个口头协议，一方面，我一定按联合国方面的要求，完成讨论会的全部工作，另一方面，我也希望会议结束后，所里不要将结余的经费挪作他用，而要用它来购置一台计算机。在我按所里的要求完成了讨论会的所有工作后，陆学艺所长也兑现了自己的承诺。于是我用剩余的会议经费 2000 美金，购置了所里第一台 IBM 系统的 PC。这是一台美国华裔生产的 sun286，有了这台机器，我才有可能开展以后的中国百县市调查和七城市家庭调查。那时五城市调查已经结束多年，并不需要用它处理五城市的调查数据。但是它使我萌发了将五城市调查原来保存的 CROMCO Z80 系列的 ASCII 码数据转换成 IBM 系列的 ASCII 码数据的想法，而且立即付诸行动。我请人用这台机器模拟打印机，将从 BCMIII 上读出的 CROMCO 系列的数据，存入这台机器，完成了将五城市调查数据从 CROMCO 系统

到 IBM 系统的转换。现在想起来，这次转换是很及时的，因为不久之后机器的更新换代越来越快，这两台机器很快就不敷使用，为功能更为强大的新微机取代而被弃之不用。这就是说，如果当时未能及时将这些 CROMCO 系列的文件转成 IBM 系列的文件，那么今天重新整理这批数据就会大费周章了，真是万幸。

正是在那一年，原来五城市项目的刘金云先生从密西根大学回国探亲，给我带来了一套 DOS 版的 SYSTAT，这是我使用的第一套数据分析软件。之前我曾经试着在 SUN 上安装 DOS 版的 SPSS，但是由于 SUN 的系统与 IBM 并不完全兼容，DOS 版的 SPSS 无法在 SUN 系统中正常运行。而 SYSTAT 则可以在 SUN 系统正常运行。我立即用它生成了一套 SYSTAT 格式五城市调查数据，并与项目组原来的成员李东山合作编辑了一本《中国城市家庭——五城市家庭调查双变量和三变量资料汇编》（社会科学文献出版社，1991 年 5 月，北京）。该书是对 1984 年出版的那本中国城市家庭资料汇编部分的补充，书中收录的都是双变量和三变量的交互表。。

2019 年，我终于完成与重庆大学出版社合约的最后两本书，《社会科学研究方法百科全书》和《数据分析》的翻译，休息了一段时间后开始着手整理几十年来自己参与和主持的十几个抽样调查的数据。1992 年在与芝加哥大学和杜克大学合作进行《中国城乡居民家庭生活调查》之后，由于合作者芝加哥大学白威廉的推荐，我开始使用 STATA。我记得当时白威廉先生给我寄来的第一套 STATA 是 2.0，还是 DOS 版的。而在开始做这个工作的时候，最新的版本已经是 STATA16 了，现在已经是 STATA17 了。从 STATA14 之后的各个版本，除了数据编辑、分析和绘图功能有很大的改进之外，系统采用了新的 UNICODE 码，因此采用 stata14 以前各种版本生成的系统文件的中文标签，在 stata14 之后的各个版本中都变成了乱码，无法辨认。而当时我手头有的 STATA 是 12，还是 10 年前与香港理工大学合作进行手机调查时，理工大学的罗沛霖先生受我影响从 SPSS 改用 STATA，购买赠送给我的。换言之，我对历年调查数据的编辑整理都是用 STATA16 进行的。当然五城市数据也不例外。编辑整理的过程是，先分调查点将对应于原始调查三个卡的 ASCII 码数据转换成 STATA 数据文件，然后分调查点将三个 STATA 数据文件合并成一个包括三个卡的全部数据，最后将 9 个数据合并成一个数据——五城市调查数据（1981-1983 五城市婚姻家庭调查.dta）。借此机会我也对数据做了最后的编辑整理：给数据的变量和变量值贴上了标签。这个工作年轻人做比较合适，老头子，老眼昏花了贴起来有点费劲，但是总算干成了。在这个过程中再次仔细读了读数据并做了记录。报告如下。

最终的数据文件是一个多重记录类型的数据文件，有 A、B、C 三种记录类型，其中 A 类记录是单条记录类型的数据，记录了被调查人本人的情况。本次调查共有被调查人 5,748（5,075 户）其中 北京东河沿 701（669 户）、北京团结湖（505 户）、天津尖山街（583 户）、上海张家弄 739（646 户）、上海长春街（640 户）、上海双阳路 775（719 户）、南京四福巷 629（530 户）、成都如是庵 401（349 户）和广州某居委会 528（528 户）。由于本次调查的被调查人是样本户中的全部已婚妇女，不少样本户都有一个以上已婚妇女，所以被调查人数会多于样本户数。为便于大家了解数据的结构我在被调查人数之外，也统计了样本户数，把它们列在了被调查人数后面的括弧中。现在看来，这样的设计叠床架屋，不仅有重复之嫌，也给数据处理增加了麻烦。户被调查人数如下表所示

B 类记录属于多条记录类型数据，记录了被调查人的家庭成员的情况，每一个家庭成员都有一条单独的记录共有 22, 122 条（个家庭成员），说明本次调查涉及的被调查人的家庭成员共有 22, 122 人。其中北京东河沿 2,289、北京团结湖 2,312、天津尖山街 2,408、上海张家弄 3,031、上海长春街 2,616、上海双阳路 3,087、南京四福巷 2,742、成都如是庵 1,278、广州某居委会 2,449。两类记录的拼接结果显示，两类数据的 ID 完全对应。与 B 类记录一样，C 类记录也是多条记录类型的记录，记录了被调查人的生育情况，涉及被调查人的全部孕育

过程，即每一个孩子的怀孕、生育和抚育情况。有多少个孩子就有多少条记录。全部记录共为 25,353 条，除去天津缺失的 529 条（因为最早开始的天津尖山街调查并没有生育部分的内容），共有有效记录 24,824 条（个孕育的孩子）其中北京东河沿 2,647、北京团结湖 2,429、天津尖山街缺失、上海张家弄 3,452、上海长春街 3,165、上海双阳路 3,629、南京四福巷 3,174、成都如是庵 1,618、广州某居委会 2,655。与 A 和 B 类数据拼接也正确无误。由于这一套数据不仅时间已经很久，而且中间经过多次转换。所以在拼接之前我十分担心会出现意想不到的情况，但结果却证明我的担心是多余的。在完成数据拼接之后我有对数据文件所含 125 个变量一一做了检查。对超界和异常值做了处理。拼接和检查结果说明，原数据清洁整齐质量很好。具体的整理编辑过程，都记录在附录中有关文件中了。大家可随时查阅，以对数据的质量和结构有进一步的了解。至此，或许我可以向费雷二老，和当初一起熬夜设计，挨家挨户登门调查和彻夜不眠处理数据的诸位同道交差了。虎子擒得否？请社会学界同仁指正。