

---

中国历史量化微观大数据：

李中清-康文林团队 40 年学术回顾<sup>\*</sup>

康文林 李中清

**摘要：**近 40 年来，李中清与康文林和他们研究团队成员（文中简称李康团队）通过构建和分析基于中国档案材料的个人层面的数据库，发现了许多新历史现象，推出了一系列受关注的研究成果，包括构建大规模、长时段人口数据库，重新探讨中国长期的人口现象、家庭与亲属网络和社会经济成就，并成功推翻了此前影响中国学界甚大的马尔萨斯人口论；构建包括清代、民国以至共和国时期官员、职业技术群体、教育精英的大规模个人数据库，系统理解中国社会的职业获得、教育获得与阶层交流状况，同时修订学界过往的相关认识。截至 2021 年 11 月，这些数据库收录了大约 893 万条个人的行为和生活状况信息。本文是李康团队首次对团队所有数据库构建与研究项目进行的系统回顾与整体性梳理与分析，包括介绍每个数据库的内容、结构和显著特征，并呈现从 1979 年到现在完整的历史数据构建的学术历程；针对所有数据库项目，进行归纳分析，凝练团队在中国各类人口行为与现象、家庭和家庭组织以及不平等和社会分层等方面最重要的研究发现与结论；最后总结团队研究方法的范式意义，特别是发现历史数据、建构数据库以及探索不同数据关联、影响和比较分析不同数据等团队研究工作成功的重要性及其学术启发。

**关键词：**中国，历史人口学，历史大数据，量化历史，不平等

## 一 引言

自 1979 年至今，40 多年来，李中清--康文林研究团队投入了大量努力，用中国档案材料定位、构建和分析个人层面的数据库，并产出了一系列学术成果。最初，团队研究了中国的人口行为、家庭、亲属网络和社会经济成就，并构建了跟踪个体从出生到死亡以及跨世代家庭的数据库。近年来，团队开始研究文武官员及其他教育精英、专业精英的社会出身和职业生涯，目前正在构建的数据库及相关研究课题有：18 世纪到 20 世纪初清代文职官员的职业生涯、20 世纪大学生的社会来源和教育轨迹、民国时期（1911-1949）政府官员及专业人士的资历和职业生涯、20 世纪 40 年代中期（土地改革时期）至 60 年代中期（人民公社时期）几十万中国农民家庭在乡村建设过程中的经历等。基于上述数据库，团队已经出版了 7 本学术著作并发表了 70 多篇学术论文（大部分以英文撰写）。其中有 11 本（篇）专著（论文）曾获得 13 个最佳学术奖或同等奖项——包括英文专著 4 本、论文 2 篇，中文专著 2 本、论文 3 篇。<sup>①</sup>

---

<sup>\*</sup> 本文原刊于 Cameron Campbell & James Lee. “Historical Chinese Microdata. 40 Years of Dataset Construction by the Lee-Campbell Research Group.” *Historical Life Course Studies* 9 (2020): 130 - 157. 在撰写这篇文章时，李中清和康文林分别得到了 HK RGC GRF 16602117 (Lee PI)、HK RGC GRF 16601718 (Campbell PI) 和 HK RGC GRF 16600017 (Campbell PI) 的支持。我们非常感谢李-康团队成员的评议、修正和其他反馈。全文由汪亭亭翻译，南京大学历史学院硕士研究生侯玥然和梁晨教授进行了校对与修订。  
<sup>①</sup> 请访问团队网页 <https://www.shss.ust.hk/lee-campbell-group/publications/> 查看这些获奖出版物的列表和相

本文是对这些项目的回顾和总结。正文第一部分为数据库概述，总结了它们的内容、数据组织方式和显著特征。第二部分介绍了自 1979 年李中清教授开始系统构建中国历史人口统计学微观数据以来，团队建设各数据库的完整历史。第三部分则总结了基于数据库的各项研究成果——从早期的人口行为，到后来的社会不平等和社会分层。最后，文中对团队多年来的研究经验进行了反思。

本文首次概述了李康团队的所有数据库项目，并将各项目与其研究成果结合探讨，以便向读者阐明所做工作之全貌。这些项目均强调通过归纳法发现社会现象，追求对历史细节的详细描述、对材料来源的仔细辨析、对作为研究所涉个体及家庭所处历史和社会背景的清醒认识。每个项目的最新信息可在研究团队网站获取。<sup>①</sup>

## 二 数据库

本节将介绍团队每个数据库的基本特征和当前进展。文中将数据库分成四个类别：

(1) 家庭、亲属关系和人口行为；(2) 官员；(3) 教育；(4) 职业；(5) 乡村建设。每个类别均包括若干数据库，有些已基本完成，有些仍在建设中。数据库内人口均为中国人，时间以 18、19 和 20 世纪为主。截至 2020 年 7 月，所有数据库共计收录 8,933,629 条记录，其中主要包含 1,705,780 名不同个体的行为和生活状况信息，并涉及与他们相关的其他数十万人的信息。<sup>②</sup>表 1 总结了这些数据库的内容。目前已有五个数据库和相关文档可供下载，后文将详细介绍下载链接及其他信息。

表 1 李中清--康文林研究团队数据库（截至 2022 年 4 月）

类别（首字母缩写）	记录数	人数
家庭、亲属关系和人口（CMGPD）	3,013,861	527,325
官员（CGED）	5,329,186	562,414
教育（CUSD）	414,060	339,816
乡村建设（CRRD）	116,244	215,960
职业（CPOD）	60,278	60,265
总计	8,933,629	1,705,780

这些数据库的共同之处在于力求完整地记录一个具有实质意义、定义明确的目标群体。这种数据库建设思路，有别于从一个大的目标群体中构建在统计学上有代表性的样本以供推论的思路。比起从目标群体中抽取代表性样本，通过囊括整个目标群体，可以更详细、更宏观地描述这些举足轻重的群体所牵涉的社会、政治和经济进程。这种方法对研究上层社会群体尤其重要。举例而言，贵族、官员、大学教员和精英大学毕业生，他们虽然只占

关数据库。其中 12 位作者是当前或曾经的小组成员：康文林(Cameron Campbell),陈必佳(Bijia Chen),陈爽(Shuang Chen),董浩(Dong Hao),李中清(James Z. Lee), 李兰(Lan Li), 梁晨(Chen Liang), 任韵竹(Bamboo Y. Ren), 阮丹青(Danching Ruan), 王丰(Feng Wang), 臧晓露(Emma Zang)及张浩(Hao Zhang)。其他合著者主要来自“欧亚人口和家庭史项目”，其中包括 Tommy Bengtsson、Marco Breschi、黑须里美(Satomi Kuros u)、Matteo Manfredini、Michel Oris、津谷典子(Noriko Tsuya)和 George Alter。

<sup>①</sup>李-康团队的网站介绍了进行中的项目、相关教师、学生和录入人员，并提供了公开发布的数据和文档的链接，详情参见 [https://www.shss.ust.hk/lee-campbell- Group /](https://www.shss.ust.hk/lee-campbell-Group/)。

<sup>②</sup>家庭、亲属和人口数据库中的每条记录，只包含作为该记录重点的个人的信息，其形式类似于人口普查表条目；而其余的许多数据库中的记录，除了重点个人的详细信息外，有时还列出了其他相关人员（通常为配偶、父母或其他亲属）的姓名、职业或其他信息。

总人口的一小部分，却可能是变革的重要推动者和先行者。如果只在他们中间抽取一个代表性样本进行研究，由于囊括的案例太少，就无法对他们的组成、功能和变迁进行充分、大量的分析。

以下是对这些数据库的详细介绍。家庭、亲属关系和人口数据库以地理区域和世袭地位划分。其中，两个数据集主要包含中国东北的农村人口；第三个数据集主要包含清代皇室成员，这些皇室成员几乎都居住在北京和现在的沈阳一带。教育数据库包括民国时期几乎所有主要大学和 1949 年后两所主要大学的学生，以及 1950 年代初期以前绝大多数中国留学生。职业数据库则包括 1760 年至 1912 年之间中国绝大部分文官和许多武官，以及从 1912 年清政府倒台后到 1950 年代早期的中国专业人士（几乎全部的注册会计师、卫生专业人员、工程师、全国的大学教师以及上海和北京两地的法律专业人员）。最后，乡村建设数据库包含 20 世纪 40 年代末和 60 年代初中期（土地改革时期），正在经历乡村建设的整个村庄、生产大队、公社甚至整个县内，以家庭为单位统计的个人信息。

### （一）家庭、亲属关系和人口现象数据库

受何炳棣对中国人口历史的早期研究<sup>①</sup>以及 Louis Henry、Peter Laslett 等对英国及法国社会经济史研究成果的启发<sup>②</sup>，团队在最初收集中国历史上个体层面的微观数据时，即聚焦于家庭、亲属关系和人口现象。由此最终构建的中国多世代人口数据库（China Multi-Generational Panel Dataset，以下简称 CMGPD）适用于事件史分析法，可用于检验社群和家庭背景对人口行为和社会经济成就的影响。<sup>③</sup>表 2 总结了 CMGPD 数据库的基本信息。

表 1 家庭、亲属关系和人口现象数据库（CMGPD）

数据库名称	首字母缩写	时间跨度	家庭数	记录数	人数	启动时间
辽宁	CMGPD-LN	1774-1909	209,880 <sup>a</sup>	1,513,352	266,091	1982
辽宁补充	CMGPD-LN-S	1909-2004		38,650	38,650	1999
双城	CMGPD-SC	1866-1913	156,711	1,346,826	107,551	2004
皇室	CMGPD-IL	1633-1933		115,033	115,033	1990
合计				3,013,861	527,325	

注<sup>a</sup>：指 1789-1913 年间。在 1789 年之前，CMGPD-LN 并不以家庭来区分个人，而是以我们称之为“家庭”的行政单位来区分个人

中国多世代人口数据库-辽宁（以下简称 CMGPD-LN）和中国多世代人口数据库-双城（以下简称 CMGPD-SC）基于辽宁和双城行政区内的户口登记册。这些登记册在格式和组织上类似于定期记录家庭及其成员的人口普查。家庭和个人在每期登记册中都以大致相

<sup>①</sup> Ping-ti Ho, *Studies on the Population of China, 1368 - 1953. Harvard East Asian Series 4.* (Cambridge, MA: Harvard University Press, 1959). 中文译本为（美）何炳棣《1368-1953 中国人口研究》，葛剑雄译，上海古籍出版社，1989。

<sup>②</sup> 具体可参见 Gautier, E., & Henry, L. *La population de Crulai, paroisse normande: Étude historique. Travaux et documents - Institut national d'études démographiques*, vol. 33. (Paris: Presses universitaires de France, 1958); Laslett, P., & Wall, R. *Household and Family in Past Time: Comparative Studies in the Size and Structure of the Domestic Group over the Last Three Centuries in England, France, Serbia, Japan, and colonial North America, with Further Materials from Western Europe* (New York: Cambridge University Press, 1972).

<sup>③</sup> Dong, H., Campbell, C. D., Kurosu, S., Yang, W., & Lee, J. Z. "New Sources for Comparative Social Science: Historical Population Panel Data from East Asia," *Demography* 52. 3 (2015b): 1061-1088.

---

同的顺序排列，这有助于手动连接不同时期的记录。通过连接不同年份的个人记录构建纵向数据，不仅可以研究个体的生活史，还可以研究家庭和家族的历史。由于 CMGPD-SC 和 CMGPD-LN 的内容和结构与其他国家的人口资料相似，它们被用于比较研究，其中最典型者是下面介绍的“欧亚人口和家族史项目”。

CMGPD-SC 和 CMGPD-LN 都记录了居住地、家庭成员与户主的关系、人口统计结果（包括出生、婚姻、死亡）和社会经济地位等基本指标。新进入或退出户口登记地的人很少，这使得两数据库记录的社区相对封闭，当人口事件（如出生、死亡、婚姻、迁徙）发生时，事件发生的时间会被记录下来，因此在任何特定时间点，经历事件的风险人群都是明确的。此外，每一系列户口登记册均对目标团体有完整记录。这些数据以居民家庭为组织单位，包含多代人，并记录了每个家庭成员与户主的关系，将孩子与他们的父母、祖父母和其他亲属联系在一起——这使得我们可以将个体嵌入其所属家庭乃至更大的亲属网络中，以观察他们的生涯成就在多大程度上依赖于远亲和祖先的特征。

CMGPD-LN 收录了 1749 年至 1909 年位于今辽宁省一带的 698 个社区。<sup>①</sup>2010 年公开发布的 CMGPD-LN 涵盖 29 个行政人口的 732 份人口登记册内容，该登记册每三年记录一次这些行政人口的个人和家庭信息。数据库中包括 151 万份个人记录和 209,880 份家庭记录。通过个人记录之间的联系，团队重建了多达七代 266,091 人的历史，后来也将家庭关联起来。这些数据和附件可在校际政治和社会研究联盟（Inter-university Consortium for Political and Social Research）下载，<sup>②</sup>每个数据集包括居住在不同地理区域的特定行政人口。

CMGPD-LN 收录的社区，分布在面积与荷兰相当的一大块区域内（见地图 1），域内经济、生态、地理环境类型多样，包括依靠渔业和农业的沿海社区、种植果园和旱地农业所在的内陆社区、以及以狩猎和采集作为补充的山地社区。<sup>③</sup>社区内人口包括租种国家土地的普通农民，以及向国家提供鱼类、蜂蜜、貂皮和其他商品的专业群体。与其它中国历史资料不同，该数据库的登记册完整而详细地记录了已婚和丧夫妇女。然而，与许多中国历史资料一样，登记册往往忽略早逝的孩子，尤其是女孩。

---

<sup>①</sup> 具体可参见 Lee, J. Z., Campbell, C. D., & Chen, S. (2010). China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN) 1749 - 1909. User guide. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

<sup>②</sup> 有关 CMGPD-LN 数据和文档，请访问 <https://doi.org/10.3886/ICPSR27063.v10>。在过去三年里，文档被下载 8,377 次，数据被下载 2,821 次（统计截止 2020 年 7 月 14 日，参见 <https://pcms.icpsr.umich.edu/pcms/reports/studies/27063/utilization>）。

<sup>③</sup> 定宜庄、郭松义、李中清、康文林：《辽东移民中的旗人社会历史文献、人口统计与田野调查》，上海社会科学院出版社，2004。

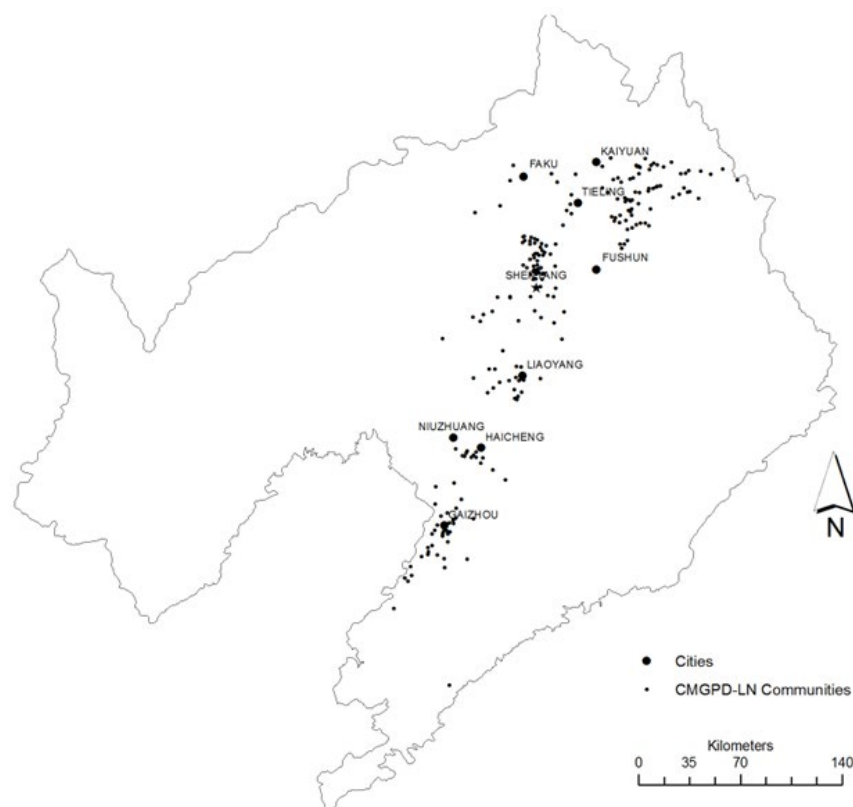


图 1 中国多世代人口数据库-辽宁的社区。

资料来源:转引自 Lee, J. Z., Campbell, C. D., & Chen, S. China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN) 1749 - 1909. User Guide (2010), p.4. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

中国多世代人口数据库-双城（以下简称 CMGPD-SC）覆盖了 1866 年至 1913 年黑龙江省双城县 125 个社区。<sup>①</sup>它包含 1,346,826 份个人记录和 156,711 份家庭记录。通过个人记录之间的联系，团队重建了多达五代、总计 107,551 人的历史，并追踪了一些家庭。图 2 展示了这些社区的地理位置，以及双城在当代中国的位置。我们于 2014<sup>②</sup>年 5 月在 ICSPR 上公开发布了 CMGPD-SC 及其相关文件。这些数据来自 14 个独立系列的登记册，每年更新登记一次。不同于覆盖辽宁东部的 CMGPD-LN，CMGPD-SC 包含的社区只在哈尔滨市正南方一个约三千平方公里的县城内。与 CMGPD-LN 相比，CMGPD-SC 的数据包含更详细的社会经济特征。例如，为了反映人口多样性，户籍记录上记录了每个家庭的官方民族身份：满族、蒙古族、汉族、锡伯族和其他民族。此外，通过将 CMGPD-SC 的家庭记录与 CMGPD-SC 的土地登记册联系起来，CMGPD 提供了每个家庭的土地财富数据，并区分了分配和自主获得的土地。和 CMGPD-LN 一样，CMGPD-SC 也详细记录了已婚和丧夫妇女。与 CMGPD-LN 相比，CMGPD-SC 虽在一定程度上记录了更多夭折儿童和女童，但

<sup>①</sup> 具体可参见 Chen, S. *State-Sponsored Inequality: The Banner System and Social Stratification in Northeast China* (Stanford: Stanford University Press, 2017); Wang, H., Chen, S., Dong, H., Noellert, M., Campbell, C. D., & Lee, J. Z. (2013). China Multi-Generational Panel Dataset, Shuangcheng (CMGPD-SC) 1866 - 1914. User Guide. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

<sup>②</sup> 在过去三年中，文件被下载 6,240 次，数据被下载 2389 次（根据 2020 年 7 月 14 日的 <https://pcms.icpsr.umich.edu/pcms/reports/studies/35292/> 利用率）。

对他们的记录仍不完整。

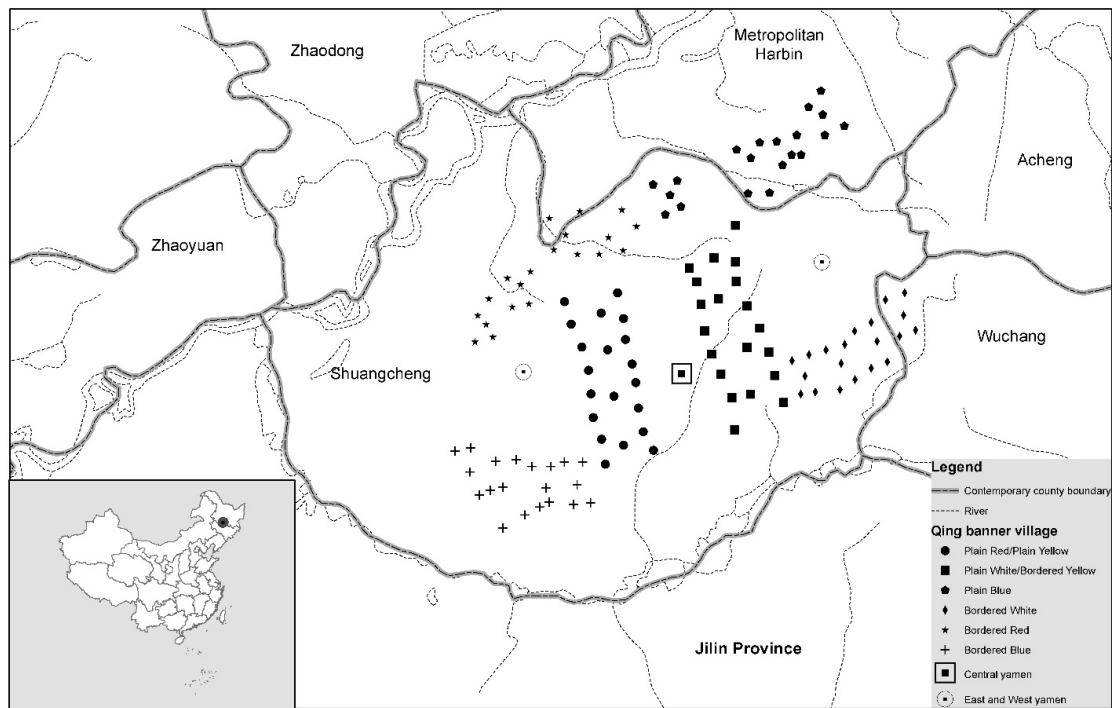


图 2 CMGPD-SC 的村庄在当代双城县的位置

资料来源：转载自 Wang, H., Chen, S., Dong, H., Noellert, M., Campbell, C. D., & Lee, J. Z.. China Multi-Generational Panel Dataset, Shuangcheng (CMGPD-SC) 1866 - 1914. User Guide. (2013), p.5. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. 原作者：Matthew Noellert。制图时使用上海交通大学历史系任玉雪提供的历史地理坐标和哈佛燕京学社中国历史地理信息系统（2007）的基础地图数据。

中国多世代人口数据库-皇室（CMGPD-IL）记录了从 1644 年（清代建立之前）到 1933 年（清代灭亡后 20 余年）115,033 名清代皇室成员及另外 135,000 名左右配偶等相关人员。<sup>①</sup>数据库内皇室成员中，约 70,000 名是宗室子女，其余约 45,000 名是旁系（觉罗）子女。<sup>②</sup>这些记录源于清代皇室《玉牒》，与中国大多数由后人所修的族谱不同，《玉牒》由宗人府专司修撰。清代宗人府有 50 至 60 名官员负责为皇室成员（主要居于北京和沈阳一带）编修《玉牒》，并掌管他们自生至死的各项事务。宗人府在 1660 年至 1921 年期间所编 28 个版本的《玉牒》，是 20 世纪中叶以前最详细完整的、关于中国大量人口生育率和婴幼儿死亡率的记录之一。与 CMGPD-LN 或 CMGPD-SC 不同，《玉牒》没有记录居民家庭构成。然而，《玉牒》记录了几乎所有皇室成员（包括女儿）的出生情况、封号和官职、重大事件以及死亡或出嫁（女儿）时间。<sup>③</sup>相比之下，那些被用在许多中国历史人口研究中的私修皇室族谱，则极少记录女儿或妻子，并且倾向于忽略在婴儿期、童年甚至青

<sup>①</sup> 蔡淑美、李中清、康文林、马文清：《“宗人府档案”数据库的建立、分析利用及其困难》，李中清、郭松义主编《清代皇族人口行为和社会环境》，北京大学出版社，1994，第 191-204 页；Lee, J. Z., Campbell, C. D., & Wang, F. “An Introduction to the Demography of the Qing Imperial Lineage, 1644 - 1911,” In R. Schofield & D. Reher, eds., *Old and New Methods in Historical Demography* (Oxford: Oxford University Press:1993):pp. 361 - 382;李中清、郭松义主编《清代皇族人口行为和社会环境》，北京大学出版社，1994。

<sup>②</sup> Wang, F., Campbell, C. D., & Lee, J. Z. “Agency, Hierarchies, and Reproduction in Northeastern China, 1789 to 1840,” In N. Tsuya, F. Wang, G. Alter, J. Z. Lee et al., *Prudence and Pressure: Reproduction and Human Agency in Europe and Asia, 1700 - 1900* (MIT Press,2010), pp. 287 - 316;汪琳岚：《清代皇族成员的先赋身份与官职获得:1644--1911》，博士学位论文，北京大学社会学系，2012。

<sup>③</sup> 鞠德源：《清代皇族人口册籍》，李中清、郭松义主编《清代皇族人口行为和社会环境》，北京大学出版社，1994，第 170-190 页。

春期夭折的儿子，以及未婚或婚后无子的成年男性。<sup>①</sup>到 19 世纪，皇室内部出现社会经济地位分化，皇帝近亲拥有各种特权，皇帝远亲则地位平凡。

## （二）教育数据库

随着研究兴趣从人口和家族史转移到社会流动和社会分层，团队扩大了数据库的收集范围，包括大学档案中的学生个人记录。这些数据库统称为中国大学生数据库（以下简称 CUSD），区分了民国时期及新中国成立后，毕业于国外大学的留学生和毕业于国内大学的大学生。这些数据主要来源于大学生的入学记录，通常包括他们的姓名、专业、籍贯、当前住址、教育背景、父母的姓名和职业，有时还包括保证人的姓名和职业。相应的数据不仅提供了学生的家庭来源信息，而且部分学生的记录可以与 CUSD 中其父母、兄弟姐妹及其他家庭成员的记录相连接。<sup>②</sup>表 3 对 CUSD 进行了总结。

表 2 教育数据库

数据库名称	首字母缩写	时间跨度	记录数	人数	启动日期
中国大学生数据库	CUSD				
民国时期	CUSD-ROC	1912-1949	165,981	136,220	2010
中华人民共和国时期	CUSD-PRC	1949-2003	150,893	150,893	1998
海外留学生	CUSD-OS	1847-1948	97,186	52,703	2015
留学日本			64,164	32,543	
留学美国			12,457	11,289	
留学欧洲			7,402	7,356	
留学苏联			766	758	
其他			115	109	
不确定			12,282	648	
总计			414,060	339,816	

中国大学生数据库-民国时期（CUSD-ROC）涵盖了民国时期（1912-1949）的大学生。<sup>③</sup>它包括 34 所民国大学的全部或部分学生注册记录。这 34 所大学只占民国时期大学总数的三分之一，但在目前能搜集到的中国大学、政府档案馆幸存的学生注册记录中，这些大学的记录已占 90%。这些大学包括大多数主要的公立、私立和教会大学。截至 2020 年 1 月，我们已经录入了这 34 所大学 136,220 名学生的 165,981 条记录。这些记录包括学生的专业、年龄、性别和籍贯。大多数学生的记录还包括至少一位家长的姓名、职业和地址，有几所学校还包括祖父母和担保人信息。父母、祖父母和担保人的信息录入工作仍在进行中，同时我们也希望从已经找到的其他大学档案中补充以上信息。

中国大学生数据库-中华人民共和国时期（CUSD-PRC）包括 1952 至 1999 年被北京大学录取的 64,500 名本科生信息，以及 1933 年至 2003 年被苏州大学录取的 86,393 名本科生信息。这些数据的重要之处，既在于它们关注中华人民共和国的精英大学生，也在于它们在时间上涵盖了 20 世纪后半叶。北京大学是中国排名最高的全国性大学之一，苏州大学

<sup>①</sup> Campbell, C. D., & Lee, J. Z. "State Views and Local Views of Population: Linking and Comparing Genealogies and Household Registers in Liaoning, 1749 - 1909," *History and Computing* 14 (1-2) (2002a, publ. 2006): 9 - 29.

<sup>②</sup> 由于录入工作正在进行中，我们预计记录的与父母或其他亲属联系起来的学生人数在未来几年将大幅增加，因此我们在这里不提供数据。

<sup>③</sup> 具体可参见 梁晨、董浩、任韵竹、李中清：《江山代有才人出——中国教育精英的来源与转变（1865—2014）》，《社会学研究》2017 年第 3 期；Bamboo Y. Ren, Chen Liang, James Z. Lee, "Meritocracy and the Making of the Chinese Academe, 1912-1952," *China Quarterly* 244 (2020): 942-968.

则是排名最高的地方大学之一。尽管自 20 世纪 80 年代以来进行的人口普查和回顾性调查中，已明确了大学毕业生的身份，但只有最近少数调查明确了他们就读的大学，因此，学界以前几乎无法研究精英院校学生群体的社会出身或地理来源，遑论追溯 20 世纪 50 年代至今如此之长的时间跨度。<sup>①</sup>

最新的中国大学生数据库-海外留学生 CUSD-Overseas (CUSD-OS) 包括 19 世纪末至 20 世纪中期在海外求学的 52,703 名中国学生，占这一时期海外留学生总数（约 6.5 万至 7 万名）的 75%至 80%。截至 2020 年 6 月,数据库共包含 32,543 名留日学生的 64,164 条记录，11,289 名留美学生的 12,457 条记录，7356 名留欧学生的 7,402 条记录，在苏联留学的 758 名学生记录，以及其它一些留学地不明的留学生记录。虽然 CUSD-OS 基于中国和外国政府记录的、在海外求学的中国学生和毕业生资料而建，几乎不涵盖这些学生的家庭成员信息，但是如果这些学生的本科学习信息也被收录于 CUSD-ROC 中，就可以找到他们的家庭成员信息。

### （三）职官及职业数据库

近年来，团队着手构建了清末文武官员职官数据库、民国和当代中国各行业者职业数据库等大型数据库。我们构建的职官及职业数据库如表 4 所示。

表 3 中国历史官员量化数据库 (CGED)

名称	首字母缩写	时间跨度	记录数	人数	启动日期
清代缙绅录	CGED-Q-JSL	1760-1911	4,433,600	327,618	2013
清代考试记录	CGED-Q-ER	1645-1907	102,484	102,484	2016
清代分发	CGED-Q-FF	1788-1903	45,337	45,337	2013
北洋职员录	CGED-BY-ZYL	1911-1924	658,873	35,814	2019
北洋分发	CGED-BY-FF	1911-1912	7,442	7,442	2019
民国	CGED-ROC	1911-1949	81,450	43,719	2018
总计			5,329,186	562,414	

职官及职业数据库中规模最大、建设最完善者是中国历史官员量化数据库--清代 (CGED-Q)。<sup>②</sup>CGED-Q 的核心资料来自于清代按季更新的官员名册——缙绅录。缙绅录列出几乎每一个实缺官员的姓名、籍贯、旗分（若有）、任职地、所任官职及其他相关信息。官员覆盖了从六部和其他中央政府部门的高品官员，到县级行政机构的低品官员。每期缙绅录记录了约 13,000 至 15,000 名官员。<sup>③</sup>截至 2021 年 11 月，我们已经录入了 1760 年 1912 年间，327,618 名官员的 4,433,600 条记录。<sup>④</sup>其中大部分记录的时间为 1830 至 1912 年，此期间留存下来的缙绅录版本比较完整。这些数据正在分阶段公布，1900 至 1912 年

<sup>①</sup> 关于 CUSD-PRC 的构建细节见梁晨、张浩、李兰等著：《无声的革命 北京大学、苏州大学学生社会来源研究 1949-2002》，生活·读书·新知三联书店，2013，第 24-37 页；关于分析方法的讨论见第 37-46 页，内容见 46-57 页。为了保护记录中学生的隐私，录入工作由大学工作人员在两所大学现场进行，原始数据未被带出。现场外的分析工作，则使用经由匿名化处理的整理、转换表或其它计算结果。

<sup>②</sup> Chen, B., Campbell, C. D., Ren, Y., & Lee, J. Z. "Big Data for the Study of Qing Officialdom: The China Government Employee Database-Qing (CGED-Q)," *The Journal of Chinese History* Volume 4, Special Issue 2 (20 20):431 - 460;任玉雪、陈必佳、郝小雯、康文林、李中清：《清代缙绅录量化数据库与官员群体研究》，《清史研究》2016 年第 4 期。

<sup>③</sup> CGED-Q 还包括一些武官名单，同样来自于最初每三个月编印一次的《中枢备览》名册。每期《中枢备览》记录了 7000-8000 名武官。

<sup>④</sup> 其中，文官记录 3,606,301 条，武官记录 518,596 条；文官出身的官员 284,264 人，武官出身的官员 6,08 07 人。



间的数据可在香港科技大学的数据空间和中国人民大学清史研究所的镜像网站下载。<sup>①</sup>

在缙绅录数据库中，可通过人名匹配方式连接同一官员在长时段内连续各季的任职记录，从而构建和研究其仕途生涯。这种连接方法的使用与该官员是否为八旗子弟有关。非八旗子弟的官员大多是汉人，可以根据他们的姓、名、籍贯省和籍贯县连接其记录。这四个变量的组合具有极高唯一性，因此连接工作面临的首要挑战，是少数情况下同一官员的姓名在两期缙绅录中可能略有不同（通常是被误记录为其他形似字或音似字），这会导致此人的不同记录没有被连接起来。八旗官员的连接则更加困难，因为他们被记录的是旗分，而非籍贯省、县。此外，大多数旗人官员都是满族或蒙古族人，没有姓氏记载。因此，连接旗人官员数据时面临的主要挑战，与连接非旗人官员（主要是汉人官员）时恰好相反。大约 86% 旗人官员的名字和旗分组合是唯一的，而对于那些名字与旗分组合存在重复的旗人官员记录，我们运用额外信息进行区分，以防混淆同一旗分的重名官员。团队目前正在录入官员的家庭背景信息、科举考试成绩记录以及其他来源的个人特征信息，并尝试建立这些信息之间的联系。

此外，团队已开始将职业数据库的覆盖范围，扩大至民国时期和中华人民共和国初期的政府官员和专业人士数据。这些数据对研究这一时期的国家建设，以及研究专业人士这一特殊社会群体的出现均有重要作用，并为对比研究清代、民国时期和新中国初期的官员和专业人士提供了可能。在中国历史官员量化数据库--民国时期（CGED-ROC）中，我们已经录入了 1911 至 1949 年间的 81,450 条政府官员记录。这些记录覆盖了来自教育部、国防部、中央研究院和国民政府五院（监察、考试、行政、司法、立法院）的 9,988 名官员，以及来自交通部和铁道部的 21,580 名官员。主要变量包括姓名、性别、年龄、籍贯、学历、当前职位和工作经历。我们还在获取相关资料，以进一步扩大该数据库。

中国专门技术人员数据库--民国时期（CPOD-ROC）（表 5）的建立可以追溯到 2016 年，当时任韵竹在 FamilySearch（前身为犹他州家谱学会）网站上找到了辽宁省档案馆的相关资料。后来，任韵竹与其他团队成员（特别是吴艺贝）合作，利用北京、杭州、南京和上海的档案馆和图书馆资料，汇编了五个独立的数据库。在目前已录入的 49,689 名专业技术人员中，18% 是医生，36% 是大学教员，36% 是工程师，剩下的 10% 是律师和注册会计师。目前，CPOD-ROC 的数据正在录入中，我们期望录入更多职业，尤其期望扩大五个在建子数据库中专业人士的规模。

表 5 中国专门技术人员数据库（CPOD）

	首字母缩写	时间跨度	记录数	人数	启动日期
民国注册会计师	CPOD-CA	1911-1949	2,021	2,008	2017
民国大学员工	CPOD-UE	1911-1949	20,340	20,340	2018
民国工程专业人士	CPOD-EP	1911-1949	20,967	20,967	2018
民国法律专业人士	CPOD-LP	1911-1949	1,350	1,350	2018
民国卫生专业人员	CPOD-HP	1911-1949	5,024	5,024	2018
中华人民共和国大学员工	CPOD-UE-PRC	1949-	5,500	5,500	2016

<sup>①</sup> 相关数据和文档于 2019 年 5 月首次在科大数据空间公开发布，详见：<https://doi.org/10.14711/dataset/E9GKRS>。截至 2020 年 7 月 14 日，香港科技大学数据空间数据下载量 741 次，李-康团队网站 CGED-Q 项目页面访问量 14099 次。我们正在与中国人民大学清史研究所合作负责发布工作，由人大清史所负责维护一个镜像站点，以便中国大陆用户访问数据和文档。截至 2020 年 5 月 7 日，该镜像网站已有 3471 名访客。

院士专家	CPOD-AE	5,076	5,076	2019
总计		60,278	60,265	

#### (四) 乡村建设数据库

最后，为了研究二十世纪中叶的中国乡村革命与乡村建设，自 2011 年起，团队在黑龙江、山西、河北等多省收集了反映农民生产生活的个人、家庭层面数据，并启动了中国乡村建设数据库（CRRD）项目。表 6 对乡村建设数据库进行了总结。

表 6 乡村建设数据库

	首字母缩写	时间跨度	记录数	人数	启动时间
中国乡村建设数据库	CRRD				
四清	CRRD-SQ	1946-1966	24,981	125,797 <sup>a</sup>	2013
土地改革	CRRD-LR	1947-1948	91,263	90,163	2011
总计			116,244	215,960	

注<sup>a</sup>：户口本只记录了 15 岁以上人员的姓名和详细信息。原始的户口表格上填写了年龄低于 15 岁的口数，但没有提供有关他们的其他详细信息。

二十世纪中叶的中国乡村建设既是近代中国社会变迁的核心，亦是世界上最大的农业社会的转型。我们正在构建两个数据集，记录 1946 年至 1966 年这一转型时期的个人和家庭经历。当时中国共产党在全国范围内进行了土地再分配，然后逐步将农村社区组织成农业合作社，最终组成人民公社。中国乡村建设数据集-土地改革（CRRD-LR）的建立是为了研究 1946 年至 1953 年的全国性土地改革运动。在这场运动中，中国许多农村地区的地方政府对土地改革中的事件和活动进行了系统记录。这些记录包括详细的个人、家庭层面的财产没收与分配记录，以及伴随着这种财富再分配的政治斗争信息。目前，CRRD-LR 包含了 1946 年至 1948 年间黑龙江双城全县 80,000 多户、约 400,000 人土地改革经历的数据。

中国乡村建设数据集--四清（CRRD-SQ）是关于 1940 年代土地改革之前到 1966 年文化大革命前夕，中国农村社会和经济变革的最系统详细的资料之一。<sup>①</sup>该数据库所采用的史料来源于 1965-1966 年社会主义教育运动期间四清运动工作队记载的《阶级成分登记表》。CRRD-SQ 目前包含山西、河北、内蒙古和广东四省 25,000 多份这样的家庭表格，其中四分之一是与山西大学中国社会历史研究中心合作整理。阶级成分登记表正面为个人家庭信息，记载了每户从约 1946 年土改前到约 1966 年的经济状况、户主的社会关系、三代家史；背面为家庭成员简况，记载了该户中 15 岁（指虚岁，按西方的标准其实是大约 13.5 岁或稍长）以上个人的社会人口信息，包括性别、年龄、民族、宗教、文化程度、职业、与户主关系，以及个人简历。

### 三 数据库构建历史

回顾过去的四十年，我们可以将团队开展的数据库建设及研究工作分为三个阶段。第一阶段为 1979 年到 1989 年。这一时期，由于资金、技术和支持人员限制，转录和分析工作进展缓慢。这一时期的工作重点是对 CMGPD-LN 的早期迭代版本进行人口统计分析。

<sup>①</sup> Xing, L., Campbell, C. D., Li, X., Noellert, M., & Lee, J. Z. "Education, Class and Assortative Marriage in Rural Shanxi, China in the Mid-twentieth Century," *Research in Social Stratification and Mobility* 66 (2020):1 - 15.

---

第二阶段为 1990 年到 2010 年。由于有了稳定的资金以支持核心工作团队，这一时段数据录入工作的速度加快了。核心工作团队由全职数据编码员组成，工作地点首先在美国，后来在中国。数据覆盖范围扩大到目前的整个 CMGPD 以及 CUSD-PRC。第三阶段为 2010 年至今。随着 CGED、CPOD、CUSD-ROC 和 CUSD-OS 的启动以及关于乡村建设主题 CRRD-LR 和 CRRD-SQ 的建立，数据库收录人员类别的范围扩及政府官员、专业人士和其他教育精英。本节将分别介绍这三个阶段的进程，尤其是每个项目的启动时间和启动方式、参与者及其贡献、以及数据库构建方法和规模的关键转变。

### （一）第一阶段--开始，1979-1989 年

四十多年前，1979 年冬天，李中清访问了北京第一历史档案馆，自此便开始在中国大陆的历史档案中寻找可量化的个人层面微观数据。在 20 世纪 60 年代和 70 年代，计量史学家和社会学家用档案构建了数据库并开展分析，改变了学界对欧洲、北美历史上家庭和人口的理解——受此启发，李中清希望以同种方法研究中国历史。<sup>①</sup>

1982 年，在鞠德源的建议下，李中清访问了辽宁省档案馆，并获得了涵盖道义县 1774 至 1798 年间的五份户口登记册的微缩胶片。<sup>②</sup>李中清与 Robert Eng（一位经济史学家，之前有处理日本历史人口登记的经验）一起制定了一个编码方案，并亲自将 1774、1780、1786 和 1792 年的登记册内容转录为固定表格格式，先用纸抄录，然后转为数字文件。<sup>③</sup>1984 年，李中清还参加了加州大学伯克利分校人口学研究生小组开设的人口统计方法课程。他与加州理工学院（Caltech）的多名本科生合作，发表了第一份利用户籍微观数据分析 20 世纪前中国大陆特定历史人口死亡率、生育率和家庭结构的成果。<sup>④</sup>他还与 William Lavelly 及王丰一起发表了一篇颇具影响力的文章，探讨新发现的历史性数据及当代微观数据如何重塑对中国人口行为的理解。<sup>⑤</sup>

虽然李中清从 1985 年起获得了更多的 19 世纪道义县的户口登记册，但对这些微观数据的研究直到 1987 年夏天康文林的加入才取得显著进展。康文林当时是加州理工学院电气工程专业的二年级学生，在高中时就对中国历史产生了兴趣。<sup>⑥</sup>他此前接受过数据库编程方面的培训，并具有相关经验。在仔细研究了为道义户口登记册设计的数据转换、计算 C 语言程序后，康文林为李中清设计了一套新的工作流程：首先使用 dBase III+ 系统（后来

---

<sup>①</sup> 这十年的资金来源既有加州理工学院的内部资源，也有国家人文基金会和 Wang Fellowship 中国研究奖学金的外部支持。此外，在北京、沈阳和台北的差旅和研究经费几乎全部来自中国国家高等学习研究计划和台湾“中央研究院”。

<sup>②</sup> 1985 年和 1987 年在辽宁省档案馆额外拍摄的登记册，将该数据涵盖时间范围延长至 1873 年。

<sup>③</sup> 请参阅 Lee & Campbell（1997, pp. xix-xxi），以了解协助道义县登记册录入和分析工作的完整人员名单，参与数据录入工作的编码人员名单，以及资金来源情况。

<sup>④</sup> 具体可参见，安醮匿、李中清、陈慧雯：《一七九五至一八二〇年间辽宁农村的成人死亡率》，中国第一历史档案馆编《明清档案与历史研究 中国第一历史档案馆六十周年纪念论文集》，中华书局，1988，第 885 - 898 页；Lee, J. Z., & Eng, R. “Population and Family History in Eighteenth-Century Manchuria: Preliminary Results from Daoyi 1774 - 1798,” *Ch'ing-shih wen-t' i* 5. 1 (1984): 1 - 55; Lee, J. Z., & Gjerde, J. “Comparative Household Morphology of Stem, Joint, and Nuclear Household Systems: Norway, China, and the United States,” *Continuity and Change* 1. 1 (1986): 89 - 111. 此时，Arthur Wolf(武雅士)已经利用 20 世纪上半叶日本殖民时期的记录，对台湾进行了相关研究，具体可参见，Wolf, A., & Huang, C.-s. *Marriage and Adoption in China, 1895 - 1945* (Stanford: Stanford University Press, 1980)。

<sup>⑤</sup> Lavelly, W., Lee, J. Z., & Wang, F. “Chinese Demography: The State of the Field,” *Journal of Asian Studies* 49. 4 (1990): 807 - 834.

<sup>⑥</sup> 康文林在 1987 年和 1988 年夏季得到了加州理工学院夏季本科生研究奖学金的支持，以供参与项目。

---

使用 dBase IV) 进行数据管理, 进而将数据导入 SPSS 软件进行分析。1987 年夏天, 与李中清共同重访北京第一历史档案馆和辽宁省档案馆后, 康文林开始着手开发新程序。<sup>①</sup>新的数据处理流程包括构建人口事件标记变量、识别不同登记册中的同一人记录、连接亲属关系、判定家庭结构及组成、以及构建人际关系网络(包括特定亲属网络的存在与否)。这些工作简化了数据输入过程, 并为数据分析创造了新的可能性, 使分析方法不再局限于比率计算。<sup>②</sup>

## (二) 第二阶段--加速, 1990-2009 年

得益于 1990 年台湾中央研究院和台湾科学委员会的支持, 以及 1999 年李中清的大量私人捐赠, CMGPD-LN 的录入速度加快。1990 年以前, 在最终构成 CMGPD-LN 的 29 个行政区域中, 只有道义的全部登记册, 以及盖州的部分登记册录入完成, 两地共计约 10 万条数据。相比之下, 1990 年至 1999 年期间的录入工作迅速得多——我们从另外 8 个行政区域中录入了约 400000 条记录。

之所以能提高录入速度, 一部分原因是犹他州家谱学会从辽宁省档案馆获得了大部分馆藏户口登记册和相关资料, 并将其对我们开放; 另一部分原因是研究资金的增加使我们能够支持一个更大的录入团队, 以便将这些资料转录为数据库。<sup>③</sup>几乎所有数据录入工作都在美国开展, 但有两个 CMGPD-LN 数据集在台湾完成。1999 年, 团队将数据录入工作地转移到了中国大陆, 在那里我们很幸运地找到了一位(不久后变成三位)可靠并热情的全职录入人员。他们分别是肖星、孙惠成和激扬。<sup>④</sup>在接下来的四年里, 他们输入了剩余的 19 个 CMGPD-LN 数据集, 为数据库增加了 100 万条数据。随后, 在 2010 年他们耗费一年时间为整个 CMGPD-LN 的 150 万条数据进行了数据清理, 为公开发布这些数据做好了准备。<sup>⑤</sup>

为了对生育率和婴幼儿死亡率进行比 CMGPD-LN 更详细的分析, 李中清与台湾中央研究院的赖慧敏和刘素芬合作, 于 1990 年开始构建 CMGPD-IL。第一历史档案馆的鞠德源此前曾向李中清介绍了宗人府收集的历史人口微观数据, 包括《玉牒》等皇室宗谱。<sup>⑥</sup>

《玉牒》几乎完整地记录了男性和女性的出生以及婴儿和儿童的死亡, 而原本的 CMGPD-LN 很少记录女儿, 并遗漏了一些在婴儿期或童年夭折的儿子, 因此《玉牒》成为对

---

<sup>①</sup>1987 年夏天, 为了在北京和沈阳进行编程和分析, 李中清和康文林从美国带来了一台又大又重的便携式微型计算机。然而, 由于忘记了变压器在中国尚未普及, 在沈阳工作时, 不得不依靠一个手工制造的闸刀开关变压器。

<sup>②</sup>1989 年 6 月康文林毕业时, 他已经从修习电气工程转修历史与工程及应用科学双学位。毕业后, 他在 Watson 和 Durfee 基金会的资助下前往台北和北京学习中文, 并在道义县生活了一段时间。在台北, 他花了六个月的时间在台湾师范大学的普通话培训中心集中学习中文。1990 年夏天, 他在台北与李中清一起工作。1990 年秋, 他前往宾夕法尼亚大学攻读人口学博士学位。1992 年至 1993 年, 他又在地处台北的台湾大学的校际中文研究项目(Inter-University Program in Chinese Language Studies)进行了一年的中文强化训练, 并于 1993 年夏天在北京进行了论文课题研究。此间他获得了社会科学研究理事会国际论文研究奖学金(Social Science Research Council International Pre-Dissertation Research Fellowship)的支持。

<sup>③</sup>在 CMGPD-LN 完成后, 我们向犹他州家谱学会(现为 FamilySearch)提供了一份数据库记录与原始资料页码的对照表, 以便纳入他们的在线搜索系统。

<sup>④</sup>将数据输入工作转移到中国后, 另一个后勤方面的挑战是给编码员发放工资。我们非常感谢美国密歇根大学 ICPSR、加州大学洛杉矶分校 CCPR 和香港科大的工作人员对这些资助的监督和管理, 特别是 ICPSR 的 Ruth Danner、加州大学洛杉矶分校 CCPR 的 Lucy Shao 和香港科大的 Freda Ching。

<sup>⑤</sup>1999 年至 2006 年, 我们还走访了辽宁的 57 个村庄, 整理了近 250 个家谱等相关数据集。

<sup>⑥</sup>鞠德源:《清代皇族人口册籍》, 李中清、郭松义主编《清代皇族人口行为和社会环境》, 北京大学出版社, 1994, 第 170-190 页。

CMGPD-LN 的宝贵补充。1985 年北京第一历史档案馆为犹他州家谱协会拍摄《玉牒》的缩微胶片时，李中清获得了这些数据的副本。1990 年至 1992 年间，李中清与赖慧敏、刘素芬合作，监督了针对宗亲的录入工作。李中清还在 1989 年招募了王丰参与这些数据的分析，并与康文林一起发表了 CMGPD-IL 数据集的介绍文章。<sup>①</sup>康文林在他的博士论文中使用了 CMGPD-IL。他通过比较 CMGPD-IL 中 18 和 19 世纪的死亡模式，与北京在 20 世纪 20 和 30 年代以及 1949 年后的死亡模式，来研究北京的长期死亡趋势。<sup>②</sup>后来，李中清与其博士生汪琳岚（北京大学社会学系）合作，在数据库中加入 1933 年《爱新觉罗族谱》中记录的旁系子弟（觉罗）数据，并将其运用于汪琳岚的博士论文写作。<sup>③</sup>

2003 年，当时就职于密歇根大学的李中清与历史学博士生陈爽合作，开始构建 CMGPD-SC，以使用黑龙江双城县土地持有和人口行为的关系问题。<sup>④</sup>除了土地持有，CMGPD-SC 还可以进行 CMGPD-LN 无法实现的其他分析，例如基于民族的比较分析等。犹他州家谱学会共获得了双城 338 份人口登记册、23 份土地登记册，并于 2003 年秋季将其向我们开放。2004 至 2007 年，肖星、孙惠成和激扬录入了 CMGPD-SC 的 130 万条记录。<sup>⑤</sup>相比之下，CMGPD-LN 的 150 万条记录的录入则进展缓慢，大约花了 20 年时间。现任爱荷华大学历史系副教授的陈爽对数据录入工作进行了监督，领导了对 CMGPD-SC 的分析工作，并以此作为她博士学位论文和专著的主要内容。<sup>⑥</sup>这同时也是我们对“欧亚项目-生育卷”的主要贡献。<sup>⑦</sup>

由于工作流程的一些变化，团队稳定管理和分析大型数据库的能力得到了提高。20 世纪 90 年代初，团队使用由马文清（Chris Myers）和康文林维护 dBase 程序。这些程序读取录入人员提供的原始录入文件，并生成可以在 SPSS 和后来的 STATA 中进行分析的文件。但是 dBase 程序工作速度很慢。20 世纪 80 年代末，需要处理的记录接近 70,000 条时，为了将录入人员输入的原始数据转化为可用于分析的文件，dBase 程序需要运行一天以上，

<sup>①</sup>具体可参见 Lee, J. Z., Campbell, C. D., & Wang, F., "An Introduction to the Demography of the Qing Imperial lineage, 1644 - 1911," In R. Schofield & D. Reher, eds., *Old and New Methods in Historical Demography* (Oxford: Oxford University Press, 1993), p. 361. 书中列出了推动这项工作的中央研究院的学者，输入数据的编码员，帮助编程的加州理工学院的本科生，以及资金来源。李中清、郭松义主编的《清代皇族人口行为和社会环境》一书提供了该项目详细的中文历史，包括数据输入过程和为分析做准备的数据转换。

<sup>②</sup> 1993 年夏天，康文林在北京做档案和图书馆研究时，收集了 20 世纪北京的死亡率数据。这项研究得到了黄志环(Jennifer Huang Bouey)的协助。

<sup>③</sup> 与《玉牒》不同，《爱新觉罗家谱》中国历史上几乎所有家谱一样，不涉及女儿。2008 年 4 月至 2011 年 7 月，数据录入工作主要由激扬在肖星的协助下完成。汪琳岚：《清代皇族成员的先赋身份与官职获得：1644--1911》，博士学位论文，北京大学社会学系，2012；Wang, L., Lee, J. Z., & Campbell, C. D. "Institutions and Inequality: Comparing the Zongshi and the Jueluo in the Qing Imperial Lineage," *Sungkyun Journal of East Asian Studies* 10. 1 (2010): 33-61.

<sup>④</sup> 2013 年公布的 CMGPD-SC 用户指南提供了一份以各种方式为 CMGPD-SC 构建做出贡献的完整个人名单，具体可参见，Wang, H., Chen, S., Dong, H., Noellert, M., Campbell, C. D., & Lee, J. Z. (2013). *China Multi-Generational Panel Dataset, Shuangcheng (CMGPD-SC) 1866 - 1914. User Guide*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

<sup>⑤</sup> 此期间的工作得到了 NIH NICHD 1R01HD045695-01A2 的支持（Demographic Responses to Community and Family Context, James Lee PI）。

<sup>⑥</sup> Chen, S. *Where Urban Migrants Meet Rural Settlers: State Categories, Social Boundaries, and Wealth Stratification in Northeast China, 1815-1913* (Ph. D. diss., University of Michigan, Department of History, 2009); Chen, S. *State-Sponsored Inequality: The Banner System and Social Stratification in Northeast China* (Stanford: Stanford University Press, 2017).

<sup>⑦</sup> Chen, S., Campbell, C. D., & Lee, J. Z., "Categorical Inequality and Gender Difference: Marriage and Remarriage in Northeast China, 1749 - 1913," In C. Lundh, S. Kurosu et al, eds., *Similarity in difference: Marriage in Europe and Asia, 1700 - 1900* (Cambridge, MA: MIT Press, 2014), pp. 393 - 438.

---

而且很容易崩溃。<sup>①</sup>最后，在 90 年代中期，我们决定停止继续开发 dBase 程序。该程序继续被用来处理传入文件，并为 STATA 分析准备文件，但没有被进一步开发。新变量创建改为在 STATA 中进行。最终，康文林决定完全放弃使用 dBase 程序，并编写了 STATA 代码来处理导入文件、数据运行、新变量创建和数据分析、数据输出的整个过程。这使得将录入人员提供的原始文件转化为分析用的工作文件所需时间，减少到几个小时。

为了更好地了解 CMGPD-LN 记录的社会背景，并更多地了解记录中家庭的历史，我们在辽宁农村开展了实地调查。1999 年至 2006 年期间，我们在辽宁省及地方志办公室的高静及其同事的陪同下，对辽宁进行了 8 次实地考察，访问了 57 个大型农村社区。我们花了大约 250 工日走访 CMGPD-LN 人口的后代，并收集当地的资料，如家谱、墓碑、契约和其他关于这些人口的家传资料。我们还收集了从 1911 年（CMGPD-LN 截止日期）到我们访问时的口述历史和家庭信息。我们将这些地方数据与康文林和李中清以及定宜庄、郭松义、李中清和康文林研究的州家庭和人口登记册进行了比较。<sup>②</sup>在每个社区，我们与受访家庭分享了从 CMGPD-LN 中生成的家谱。许多受访家庭已经遗失了他们的家谱，或仅存列出男性成员辈分和名字的简易家谱。我们提供的材料，包括了他们曾出仕的、取得一定成就的、或拥有一定知名度的祖先姓名等信息，这帮助他们重建了家族历史。

得益于技术进步，我们开始使用事件史分析和其他回归分析方法，来研究个人人口行为和结果与家庭和社区背景之间的关联。用于处理和分析数据的个人电脑，性能飞速提升。在 20 世纪 90 年代初，涉及道义数据集中 10 万条左右记录的计算，除了制表或线性回归外，其他任何处理都需要 15 分钟到 1 小时，所需具体时长取决于所包括的样本数、变量数、模型类型和模型数量。到 20 世纪 90 年代末，在更大数量的记录上运行更高级的运算，所需时间则要少得多。到 2010 年，对 CMPGD-LN 和 CMGPD-SC 近 300 万条记录的合并计算，几分钟内即可在个人电脑上完成。

1993 年，速水融 (Akira Hayami) 邀请李中清参加“欧亚大陆人口和家庭史项目”。这是一个国际比较项目，研究古代社区背景、家庭组织和人口行为之间的相互作用，这也激励了我们向事件史分析的转变。<sup>③</sup>该项目于 1994 年启动，在李中清、Bengtsson 和 Alter 的编

---

<sup>①</sup> 当康文林还在加州理工学院时，电脑程序非常慢，而且容易崩溃。因此当他和李中清向数据库合并新编码的登记册、或向数据库增加新变量时，他经常睡在李中清办公室的地板上，每隔几个小时就会起来检查程序是否还在运行，一旦报错就纠正问题并重新启动程序。

<sup>②</sup> Campbell, C. D., & Lee, J. Z. “State Views and Local Views of Population: Linking and Comparing Genealogies and Household Registers in Liaoning, 1749 - 1909,” *History and Computing* 14 (1-2) (2002a, publ. 2006): 9 - 29; 定宜庄、郭松义、李中清、康文林：《辽东移民中的旗人社会 历史文献、人口统计与田野调查》，上海社会科学院出版社，2004。

<sup>③</sup> 各卷的历史可参见 Bengtsson, T., Campbell, C. D., Lee, J. Z. et al, *Life under Pressure: Mortality and Living Standards in Europe and Asia, 1700 - 1900*(Cambridge, MA: MIT Press,2004),pp.xi-xii;Lundh, C., Kurosu, S. et al, *Similarity in difference: Marriage in Europe and Asia, 1700 - 1900*,( Cambridge, MA: MIT Press,2014),p p.xxiii-xxv,xvii-xxi;Tsuya, N.O., Wang, F., Alter, G., Lee, J. Z. et al, *Prudence and Pressure: Reproduction and Human Agency in Europe and Asia, 1700 - 1900* ( Cambridge, MA: MIT Press,2010),pp. xxiii-xxv。 *Life under Pressure* 的中译本为（瑞典）托米·本特森，（美）康文林，李中清等《压力下的生活：1700-1900年欧洲与亚洲的死亡率和生活水平》，社会科学文献出版社，2007。康文林对参与该项目的经历进行了回忆和反思，可参见 Campbell, C. D. , “Large-Scale Collaboration and Comparison in Historical Demography: Reflections on the Eurasia Project,” In K. Matthijs, S. Hin, J. Kok, & H. Matsuo (Eds.), *The Future of Historical Demography: Upside and Inside Out* (Leuven: Acco,2016),pp. 193 - 196;Lee 和 Steckel 对 Bengtsson 等人的工作表达了欣赏和评价，可参见 Lee, R., & Steckel, R. H. “Life under Pressure: An Appreciation and Appraisal,” *Historical Methods* 39. 4 (2006):171 - 176. Retrieved from [https://u.demog.berkeley.edu/~rlee/oldceda/Publications/pdfs/ rlee/LeeSteckel\\_LifeUnderPressureReview\\_HistoricalMethods2006.pdf](https://u.demog.berkeley.edu/~rlee/oldceda/Publications/pdfs/ rlee/LeeSteckel_LifeUnderPressureReview_HistoricalMethods2006.pdf)。 Daniel Little 在他的博

---

辑下，麻省理工学院出版社出版了该系列的三本专著。第一本关于死亡率，第二本关于生育率，第三本关于婚姻。<sup>①</sup>该项目由拥有比利时、中国、意大利、日本和瑞典社区户籍数据的研究团队合作，指定了死亡率、生育率和婚姻的事件史模型——这些模型可以应用于所有数据集中，从而产生可比较的结果。后文将讨论该项目的发现。

21 世纪初，团队开始计划公开发布 CMGPD。李中清在 2003 年到密歇根大学工作以后，加入了校际政治和社会研究联盟（ICPSR），这对我们的工作至关重要。与 Myron Guttman 等人的谈话使我们相信，公开发布 CMGPD 不仅重要，而且在 ICPSR 的支持下也是完全可行的。ICPSR 为我们向美国国立卫生研究院提交资金申请和管理以及拨款管理提供了行政支持，并通过 ICPSR 的人口研究数据共享（DSDR）计划为 CMGPD 的数据、文档存放提供了人员支持。<sup>②</sup>2009 年李中清调到香港科技大学工作（HKUST）后，加州大学洛杉矶分校（UCLA）的康文林在加州人口研究中心的行政支持下，负责提交和管理一项拨款，以支持 CMGPD-SC 的公开发布。但公开发布工作本身仍然通过 ICPSR DSDR 进行，并得到了 ICPSR 工作人员的支持，其中最主要的工作人员是 Susan Hautaniemi Leonard。从 2011 年到 2014 年，康文林每年夏天在上海交通大学举办讲习班，向用户介绍 CMGPD 的内容和数据组织方式，并演示管理和分析数据的高级操作。在此期间由任玉雪接待康文林，Dan Xu 提供后勤支持。基于讲习班的培训经验，康文林、董浩和李中清编写了《CMGPD 培训指南》，作为《CMGPD 用户指南》的补充。<sup>③</sup>

2000 年代末期，团队在社会流动、分层和不平等方面开辟了新研究路线。最初，我们研究了道义数据库中父子之间社会和经济成就方面的关联（Lee & Campbell 1997:196-214）。1996 年康文林刚开始担任加州大学洛杉矶分校社会学助理教授时，仍然专注于人口学，但与 Donald Treiman、William Mason、Robert Mare、Judith Seltzer、Ken Sokoloff、Jean-Laurent Rosenthal 以及其他社会学和经济学同事的交流，激发了他利用 CMGPD-LN 的独特属性研究社会流动、不平等、亲属关系和其他主题的想法。<sup>④</sup>除了研究父子之间的关联外，我们进一步研究了社会经济成就与扩展的亲属网络特征之间的关联——从同一家

---

客上发表了文章《理解社会》，讨论了“欧亚人口与家庭史项目”，参见：<https://understandingsociety.blogspot.com/2014/08/eurasia-project-on-population-and.html>。

<sup>①</sup> 我们还要感谢在 2004 年、2010 年和 2015 年的社会科学史协会会议，以及 2005 年和 2011 年的美国人口协会会议上评议这些作品提学者们，以及 2014 年美国社会学协会会议对“欧亚人口与家庭史项目”的总结讨论。在这些会议的众多参与者中，有 Douglas Anderton, Jason Beckfield, Hilde Bras, Andrew Cherlin, Jack Goldstone, David Hacker, Michael Haines, Charles Hirschman, Jan Kok, Ronald Lee, Daniel Little, Deirdre McClosky, Myron Gutmann, Karen Oppenheim Mason, Richard L. Steckel, Jan Van Bavel 和 Andreas Wimmer。在每一卷出版之前，我们在社会科学史协会会议、欧洲社会科学史会议、世界经济史大会和其他会议上分享了正在进行的工作。在各卷的致谢中均已列出这些会议。

<sup>②</sup> NIH NICHD 1R01HD057175-01A 项目（The Liaoning Multi-Generational Panel Dataset: Public Release and User Training, Lee and then Leonard PI）支持了 CMGPD-LN 在 2009 年至 2012 年期间的公开发布。NIH NICHD 1R01HD070985-01（Multi-generational Demographic and Landholding Data: CMGPD-SC Public Release. Campbell PI）支持了 2012 年至 2016 年期间 CMGPD-SC 的公开发布。

<sup>③</sup> Lee, J. Z., Campbell, C. D., & Chen, S. (2010). China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN) 1749 - 1909. User guide. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.; Wang, H., Chen, S., Dong, H., Noellert, M., Campbell, C. D., & Lee, J. Z. (2013). China Multi-Generational Panel Dataset, Shuangcheng (CMGPD-SC) 1866 - 1914. User Guide. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

<sup>④</sup> 在 1990 年代和 2000 年代，在加州大学洛杉矶分校举办的三个系列研讨会给了康文林关于超越传统的历史人口学方面特别重要的启发。这三个研讨会是：Von Gremp Workshop in Economic History、the UCLA/RAND Joint Labour and Population Workshop 和 the California Centre for Population Research Seminar Series。

---

庭的兄弟姐妹、叔叔和祖父母起步，然后扩及家庭外的亲属，最后扩展至世系。然而，我们面临一个限制因素：CMGPD 数据集没有记录除了成年男子的官职之外，任何关于其他非农业职业的信息。此外，只有 CMGPD-SC 记录了土地持有情况。

由于对不平等和社会流动领域的研究兴趣日益增长，并渴望跳出人口统计学领域，在花了 20 多年时间收集 20 世纪前中国社会经济成就、相关人口现象的个人层面信息之后，我们将注意力转向构建数据库以研究古今中国的不平等、社会流动和社会变革。在知晓北京大学档案馆所保存了学生学籍卡后，李中清及其合作者启动了 CUSD-PRC 项目。北京大学学籍卡录入工作启动于 2003 年，由李中清的合作者阮丹青、杨善华以及杨的博士生张浩监督。苏州大学学籍卡的录入工作则于 2007 年开始，由清华大学历史学博士梁晨监督，他在密歇根大学做过李中清的博士后，现在是南京大学历史学教授。

### （三）第三阶段——扩大，2009 年至今

这一阶段始于 2009 年李中清转到香港科技大学。2013 年，康文林也从加州大学洛杉矶分校转到香港科技大学，加入了李中清研究团队。因地理位置之便，团队得以经常去往大陆，与研究人员交流、展示成果，并探索古今中国的系统性新材料。此间我们发现了一批可以转化为数据库的新史料，并由此将研究对象由家庭组织和人口行为扩展到教育精英。慷慨而稳定的校内外资助也使得录入工作的继续扩展。<sup>①</sup>与在美国时相比，李中清和康文林在香港更容易招募、支持有相关兴趣的研究生，并与他们开展合作。自李中清在港科大工作以来，团队培养了倪志宏（Matthew Noellert）、董浩和陈必佳，截至 2020 年夏天（本文英文版完成时），他们分别是一桥大学经济学研究科的副教授、北京大学光华管理学院社会研究中心的助理教授和中国人民大学历史系的博士后。当时团队的博士生有香港科技大学的李湘宁和任韵竹、华中师范大学的薛勤、以及上海交通大学的杨莉和吴艺贝。

团队建设了相关数据库，以研究古今中国的不平等、社会流动和社会变化等问题。2010 年，梁晨提出了一项关于 20 世纪上半叶大学生社会来源的研究计划，该研究拟在 CUSD-PRC 的基础上进行扩展，构建并分析基于中国各地历史档案馆藏学籍卡的数据库（CUSD-ROC）。<sup>②</sup>他与李-康研究团队的其他成员合作，找到了目前 CUSD-ROC34 所大学中一半学校的学籍卡，并组织了大部分的数据录入工作。任韵竹、李中清和张铭雨找到了另一半学籍卡并完成了录入工作。这些材料的转录过程与 CMGPD 数据库的转录过程不同：不是由一个专门团队负责转录原始资料的扫描件，而是在当地招聘人员到档案馆现场输入数据。因此，增加变量或因前后信息不一致核查原件时，都需要实地往返档案馆。

团队启动的下一个项目是 CGED-Q。2013 年，密歇根大学博士后、现上海交通大学历史系副教授任玉雪向康文林和李中清展示了她从清华大学图书馆出版的 206 种《缙绅录》中转录的中国东北官员记录，以及正在开展的相关研究。<sup>③</sup>康文林由此设想将《缙绅

---

<sup>①</sup>研究得到了香港研究资助局以下基金支持：642911（Lee PI）、640613（Lee PI）、16400114（Campbell PI）、16400714（Lee PI）、16602315（Lee PI）、16600017（Campbell PI）、16602117（Lee PI）、16601718（Campbell PI）。

<sup>②</sup>具体可参见梁晨、董浩、任韵竹、李中清：《江山代有才人出——中国教育精英的来源与转变（1865—2014）》，《社会学研究》2017 年第 3 期；Bamboo Y. Ren, Chen Liang, James Z. Lee, “Meritocracy and the Making of the Chinese Academe, 1912-1952,” *China Quarterly* 244 (2020):942-968.

<sup>③</sup>英文介绍参见 Chen, B., Campbell, C. D., Ren, Y., & Lee, J. Z. “Big Data for the Study of Qing Officialdom:



---

录》作为研究清代官场和清代官员职业动态的一种资源。康、李和任三人制定了一个计划，将这批材料中的 280 万条记录以及来自其他《缙绅录》版本的 120 万条记录全部录入。这项工作已于 2020 年夏天完成。2014 年 CMGPD 录入人员开始输入数据，2016 年增加了新录入人员，录入的速度提高了一倍。陈必佳于 2015 年加入该项目，当时她是香港科技大学社会科学专业的硕士生。她在数据录入的协调工作中发挥了关键作用，并撰写了关于清代官员职业生涯的博士学位论文。<sup>①</sup>

虽然 CMGPD、CGED 和大部分 CUSD 数据库都由李中清或康文林发起，并与我们研究团队的其他资深成员（如梁晨）合作完成，但我们最新关于中国农村革命（CRRD-LR、CRRD-SQ）和中国专业人士兴起（CPOD）的数据库，则主要由年轻团队成员发起——他们发掘了相关材料，并为完成博士学位论文构建了数据库。2011 年，倪志宏在双城进行实地考察时发现了 CRRD-LR 的基础材料，并利用这些数据撰写了他 2014 年的博士学位论文和 2020 年的著作。<sup>②</sup>山西大学中国社会史研究中心主任行龙根据他收集的 7,800 份农村家庭社会阶层登记表，发起了中国乡村建设--四清数据库（CRRD-SQ）。2015 年，倪志宏与山西大学中国社会史研究中心教授胡英泽合作负责了 CRRD-SQ 的初始录入工作。<sup>③</sup>2016 至 2019 年期间，倪志宏与香港科技大学博士生李湘宁合作，将 CRRD-SQ 的涵盖范围从 1 个省扩展到 4 个省，从 8000 个家庭扩展到 2.5 万个家庭。<sup>④</sup>

同样，虽然 CUSD-ROC 和 CUSD-OS 分别由梁晨在 2010 年、李中清在 2019 年发起，但从 2018 年开始，是由任韵竹发起并协调各种 CPOD 数据集的建设工作。在任韵竹的指导下，吴艺贝和杨莉协调了 CUSD-OS 的数据录入工作。吴艺贝还找到了 CGED-ROC 的数据资料，并负责了数据转录工作。

## 四 研究发现

本节将按主题来分别介绍团队的研究发现。首先，我们将介绍关于人口现象和家庭组织的研究。对该主题的研究是一个发展的过程，肇始于计算人口比率和描述家庭结构，后深入至研究因家庭背景的差异而产生的家庭等级，最后扩展至研究婚配选择、家庭背景对个人健康和晚年死亡率的影响以及其他相关主题。第二，我们将介绍关于代际社会流动和不平等的研究。我们先是研究父亲的社会经济成就如何影响儿子取得的相关成就，后来又研究父亲以外的多代亲属在塑造个人成就中的作用。最近我们把亲属网络和世系群体（而

---

The China Government Employee Database-Qing (CGED-Q),” *The Journal of Chinese History* Volume 4, Special Issue 2 (2020):431 - 460, 中文介绍参见任玉雪、陈必佳、郝小雯、康文林、李中清：《清代缙绅录量化数据库与官员群体研究》，《清史研究》2016 年第 4 期。内容包括项目介绍、结果样本以及为 CGED-Q 做出贡献的合作者、学生和录入人员的完整名单。

<sup>①</sup> Chen, B. *Origins and Career Patterns of the Qing Government Officials (1850–1912): Evidence from the China Government Employee Dataset-Qing (CGED-Q)* (Ph. D. diss., Hong Kong University of Science and Technology Division of Social Science, 2019).

<sup>②</sup> 孙惠成、肖星和激扬在 2012 年完成了 CRRD-LR 大部分的编码工作，并在 2013 年将各 CRRD-LR 事件登事件登记册上的个人联系起来。

<sup>③</sup> 由此产生的数据库保存在山西大学，分析工作由该校人员进行。与 CUSD-PRC 一样，对该数据库的任何分析都是基于应研究需要制作的表格或其他计算结果，而不是基于原始数据。李湘宁牵头负责这些材料的分析工作。

<sup>④</sup> 倪志宏 (Matt Noellert) 和李湘宁 (Xiangning Li) 牵头负责这些材料的分析工作。编码工作主要由李秀兰负责。

不是个人)作为分析变量,来更宽泛地研究社会分层和不平等问题。第三,我们将总结近期关于20世纪中国大学生的地理和社会来源的研究成果。最后,我们将概述最近关于清代政府官员职业生涯的研究成果。

## (一) 人口行为

在人口行为方面,团队最早的研究路径是使用总体比率来呈现死亡率、生育率、人口年龄及性别构成、家庭结构等的演化趋势和模式。李中清和 Eng 介绍了基于1774年至1798年间道义县五份户口登记册的相关数据,并描述了该地的出生率、死亡率、人口年龄构成和家庭结构。<sup>①</sup>他们在其他研究中证实,这些资料完全记录了成年男性和已婚及丧偶女性,但遗漏了许多在婴儿期或幼儿期死亡的儿子和大多数女儿。李中清、康文林和王丰介绍了 CMGPD-IL 数据库,并展示了皇室成员死亡的时间趋势和年龄模式。<sup>②</sup>李中清、康文林、安醯匿(Lawrence Anthony)和陈慧雯(Suen Chen)的研究表明,道义的死亡率水平和模式与历史上世界其他地区相似。<sup>③</sup>李中清和 Gjerde 比较了道义的家庭结构与挪威、美国的家庭结构,指出现有家庭结构分类方案的缺陷,并提出了一个更有利于中外社会比较的新分类方案。<sup>④</sup>对 CMGPD-IL 与 CMGPD-LN 的比较表明,皇室成年男性的死亡率高于辽宁道义农村地区的成年男性,这可能是因为被限制在北京生活使他们受到了“城市惩罚(urban penalty)”。这些早期研究成果,启发了一系列充分运用个人细节数据的后续研究:如探索生育率、死亡率以及其他人口现象与各种社会、经济变量的关系等。下文中将分别介绍这些研究。

### 1. 生育率

关于生育模式的早期研究结果启发了后续工作——研究节制生育对降低婚内生育水平的作用。王丰、李中清和康文林的相关研究表明,由道义和清朝皇室人口数据计算得到的婚内生育率低于欧洲,从结婚到首次生育的时间间隔、此后的生育间隔均比欧洲长得多,而且停止生育的时间也比欧洲早得多。<sup>⑤</sup>他们认为,这些现象是主动而非被动节制生育的结果。基于这些发现,李中清和王丰提出了与马尔萨斯及其继任者截然不同的观点:对生育率的预防性抑制,对20世纪之前的中国人口动态产生了重要影响;<sup>⑥</sup>20世纪中国大陆、

<sup>①</sup> Lee, J. Z., & Eng, R. "Population and Family History in Eighteenth-Century Manchuria: Preliminary Results from Daoyi 1774 - 1798," *Ch'ing-shih wen-t'i* 5. 1 (1984): 1 - 55.

<sup>②</sup> 具体可参见 Lee, J. Z., Campbell, C. D., & Wang, F. "An Introduction to the Demography of the Qing Imperial Lineage, 1644 - 1911," In R. Schofield & D. Reher, eds., *Old and New Methods in Historical Demography* (Oxford: Oxford University Press:1993):pp. 361 - 382.

<sup>③</sup> 具体可参见,安醯匿、李中清、陈慧雯:《一七九五至一八二〇年间辽宁农村的成人死亡率》,中国第一历史档案馆编《明清档案与历史研究 中国第一历史档案馆六十周年纪念论文集》中华书局,1988,第885 - 898页; Lee, J. Z., Campbell, C. D., & Anthony, L. "A century of mortality in rural Liaoning, 1774 - 1873," In S. Harrell, ed., *Chinese Historical Microdemography* (Berkeley: University of California Press,1995), pp. 163 - 182.

<sup>④</sup> Lee, J. Z., & Gjerde, J. "Comparative Household Morphology of Stem, Joint, and Nuclear Household Systems: Norway, China, and the United States," *Continuity and Change* 1. 1 (1986):89 - 111.

<sup>⑤</sup> Wang, F., Lee, J. Z., & Campbell, C. D. "Marital Fertility Control Among the Qing Nobility: Implications for Two Types of Preventive Check," *Population Studies* 49. 3 (1995):383 - 400; Lee, J. Z., & Campbell, C. D., *Fate and Fortune in Rural China: Social Organization and Population Behaviour in Rural Liaoning, 1774 - 1873* (Cambridge: Cambridge University Press,1997): pp. 83-102.

<sup>⑥</sup> 具体可参见 Lee, J. Z., & Wang, F., *One Quarter of Humanity: Malthusian Mythology and Chinese Realities, 1700 - 2000* (Cambridge, Mass.: Harvard University Press,1999). 中译本为李中清,王丰《人类的四分之一 马尔萨斯的神话与中国的现实 1700-2000》,陈卫,姚远译,生活·读书·新知三联书店,2000。

台湾和香港生育率骤降，是因为中国人继承了根据经济和其他情况主动控制生育率的历史遗存，迅速应用了限制生育的新技术。随后，我们与马尔萨斯理论支持者就中国历史人口动态的解释问题展开了激烈辩论。<sup>①</sup>康文林和李中清重新审视了生育控制问题，认为但凡适当考虑不同夫妇生育能力的差异性，就会发现明显的节育证据。<sup>②</sup>

我们在对死亡率进行研究的同时继续探索生育率差异，以阐明社区、家庭和个人环境对生育的影响。李中清和康文林比较了道义地区不同家庭结构、家中排行和社会经济地位男子的累积生子数量。<sup>③</sup>一般来说，在家庭或社会经济地位中具有特权地位的男子育有更多孩子。这些男子不仅早婚和再婚的可能性更高，而且在某些情况下，他们在婚姻中的生育率也更高。特权与生育之间呈正向关系，这与前述关于死亡率的结论形成了鲜明对比——在某些情况下，有特权的男性死亡率出人意料地更高。此外，在困难时期，即粮食价格高涨或出现气候问题时，生育率会下降。<sup>④</sup>董浩还研究了不同东亚人口中，当地家庭系统对共居亲属生育的调节作用差异。<sup>⑤</sup>其他分析主要集中在婚内生育和收养等行为上。王丰、李中清和康文林用 CMGPD-LN 重新审视了生育率，并证明生育率与经济和家庭等级制度中的地位有关。<sup>⑥</sup>康文林和李中清还研究了婚内生育率与生活在家以外的亲属特征的关联——但没有发现任何关联。<sup>⑦</sup>陈爽、李中清和康文林表明，双城的生育率与家庭土地持有量以及其他测度下的社会经济和家庭地位呈正相关。<sup>⑧</sup>王丰和李中清表明，在清代皇室家族中，多达 12.5% 的男性子嗣被由亲属收养，通常情况下，收养在保持血统连续性和实现其他目标方面发挥了重要作用。<sup>⑨</sup>

## 2.死亡率

早期对婴儿和儿童死亡率的描述性分析引出了对杀害女婴的研究，这成为李中清和王丰批判中国历史人口动态的马尔萨斯式解释的基础之一。<sup>⑩</sup>李中清和康文林等人以道义的出生和死亡登记作为间接证据，认为一些家庭杀戮或忽视女婴，进而影响了幸存儿童数量

<sup>①</sup> Campbell, C. D., Wang, F., & Lee, J. Z. "Pretransitional Fertility in China". *Population and Development Review* 28. 4 (2002):735 - 750; Lee, J. Z., Campbell, C. D., & Wang, F. "Positive Checks or Chinese Checks?" *Journal of Asian Studies* 61. 2 (2002): 591 - 607.

<sup>②</sup> Campbell, C. D., & Lee, J. Z. "Fertility Control in Historical China Revisited: New Methods for an Old Debate," *History of the Family* 15. 4 (2010b): 370 - 385.

<sup>③</sup> Lee, J. Z., & Campbell, C. D., *Fate and Fortune in Rural China: Social Organization and Population Behavior in Rural Liaoning, 1774 - 1873* (Cambridge: Cambridge University Press, 1997), pp.133-156, 177-195.

<sup>④</sup> Campbell, C. D., & Lee, J. Z. "Demographic Impacts of Climatic Fluctuations in Northeast China, 1749 - 1909," In S. Kurosu, T. Bengtsson, & C. D. Campbell, eds., *Demographic responses to economic and environmental crisis* (Kashiwa: Reitaku University Press, 2010a), pp. 107 - 132; Wang, F., Campbell, C. D., & Lee, J. Z. "Agency, Hierarchies, and Reproduction in Northeastern China, 1789 to 1840," In N. Tsuya, F. Wang, G. Alter, J. Z. Lee et al., *Prudence and Pressure: Reproduction and Human Agency in Europe and Asia, 1700 - 1900* (MIT Press, 2010), pp. 287 - 316.

<sup>⑤</sup> Dong, H. *Patriarchy, Family System and Kin Effects on Individual Demographic Behaviour Throughout the Life Course: East Asia, 1678-1945* (Ph. D. diss., Hong Kong University of Science and Technology Division of Social Science, 2016).

<sup>⑥</sup> Wang, F., Campbell, C. D., & Lee, J. Z. "Agency, Hierarchies, and Reproduction in Northeastern China, 1789 to 1840," In N. Tsuya, F. Wang, G. Alter, J. Z. Lee et al., *Prudence and Pressure: Reproduction and Human Agency in Europe and Asia, 1700 - 1900* (MIT Press, 2010), pp. 287 - 316.

<sup>⑦</sup> Campbell, C. D., & Lee, J. Z. "Long-term Mortality Consequences of Childhood Family Context in Liaoning, China, 1749 - 1909". *Social Science and Medicine* 68. 9 (2009): 1641 - 1648.

<sup>⑧</sup> Chen, S., Lee, J. Z., & Campbell, C. D. "Wealth Stratification and Reproduction in Northeast China, 1866 - 1907". *History of the Family* 15. 4 (2010): 386 - 412.

<sup>⑨</sup> Wang, F., & Lee, J. Z. "Adoption Among the Qing Nobility and Its Implications for Chinese Demographic Behaviour". *History of the Family* 3. 3 (1998):411 - 428.

<sup>⑩</sup> Lee, J. Z., & Wang, F., *One Quarter of Humanity: Malthusian Mythology and Chinese Realities, 1700 - 2000* (Cambridge, Mass.: Harvard University Press, 1999).

和性别构成。<sup>①</sup>这种杀戮和忽视行为是人们对当时的经济条件和个人生存状况的反应。如前所述，因不满足于依赖间接证据，我们建立了完整记录子女出生和死亡的 CMGPD-IL。这引出了对清朝皇室婴儿和儿童死亡率的分析，该分析提供了杀害婴儿的直接证据——女儿在出生后第一天和一个月内的死亡率异常地高，进一步表明杀婴不仅仅是经济危机或极端贫困的结果。<sup>②</sup>

另一组死亡率研究采用事件史分析法描绘死亡率差异模式，并阐明了家庭、社区和制度环境对死亡风险的影响。李中清和康文林首先提出，死亡率随社会经济地位和家中排行而变化。这种关系有时与直觉相反：男性的特权有时意味着更高的死亡风险。<sup>③</sup>死亡风险不仅取决于家庭规模和组成，也取决于家庭中是否有特定的亲属。康文林和李中清研究了家庭环境如何影响寡妇和孤儿的死亡率，并表明寡妇的死亡风险取决于她们是否有儿子。<sup>④</sup>有儿子的寡妇死亡率不受丧夫影响，但没有儿子的寡妇丧夫后死亡率会升高。董浩运用辽宁、台湾以及日本东北部三地的户籍数据，主导开展了三地家庭背景对死亡率影响的比较研究。<sup>⑤</sup>

在此基础上，团队进一步研究了经济、气候对死亡率的短期冲击，以及公共卫生干预对死亡率的长期影响。康和李根据对社会地位、家庭环境和物价相互作用的分析，认为在特权和死亡风险之间存在折衷关系，因为特权个体的死亡率对物价波动更敏感。<sup>⑥</sup>2004年，康和李使用 CMGPD-LN 中一个更大的样本，更详细地研究了死亡率水平差异和死亡率对物价波动的敏感性。男性死亡率比女性死亡率对谷物价格的波动更敏感，而且这种反应受年龄、社会经济地位和家庭背景的影响。<sup>⑦</sup>这些结果有助于 Bengtsson 等人对东西方的比较研究。<sup>⑧</sup>2010年，康和李调查了 1782-1789年、1813-1815年和 1831-1841年异常寒冷的夏天，以及其他异常气候对死亡率的影响。在 1782-1789年，预期寿命下降了 10 年以上。年轻男性和女性受到的打击尤其大，5-15岁男性的死亡率为非气候异常期的 8.78 倍，5-15岁女性的死亡率为 4.65 倍。<sup>⑨</sup>此外，康文林通过比较 19 世纪 CMGPD-IL 中死亡率和 20 世纪

<sup>①</sup> Lee, J. Z., Campbell, C. D., & Tan, G. "Infanticide and Family Planning in Rural Liaoning, 1774 - 1873," In L. Li & T. Rawski, eds., *Chinese History in Economic Perspective* (Berkeley: University of California Press, 1992), pp. 149 - 176; Lee, J. Z., & Campbell, C. D., *Fate and Fortune in Rural China: Social Organization and Population Behaviour in Rural Liaoning, 1774 - 1873* (Cambridge: Cambridge University Press, 1997), pp. 58-82.

<sup>②</sup> Lee, J. Z., Wang, F., & Campbell, C. D. "Infant and Child Mortality Among the Qing Nobility: Implications for Two Types of Positive Check," *Population Studies* 48. 3 (1994): 395 - 412.

<sup>③</sup> Lee, J. Z., & Campbell, C. D., *Fate and Fortune in Rural China: Social Organization and Population Behaviour in Rural Liaoning, 1774 - 1873* (Cambridge: Cambridge University Press, 1997), pp. 133-156, 177-195.

<sup>④</sup> Campbell, C. D., & Lee, J. Z. "When Husbands and Parents Die: Widowhood and Orphanhood in Late Imperial Liaoning, 1789 - 1909". In R. Derosas & M. Oris, Eds., *When dad dies* (Bern: Peter Lang, 2002b), pp. 313 - 334.

<sup>⑤</sup> Dong, H. *Patriarchy, Family System and Kin Effects on Individual Demographic Behaviour Throughout the Life Course: East Asia, 1678 - 1945* (Ph. D. diss., Hong Kong University of Science and Technology Division of Social Science, 2016); Dong, H., Manfredini, M., Kurosu, S., Yang, W., & Lee, J. Z. "Kin and Birth Order Effects on Male Child Mortality: Three East Asian populations, 1716 - 1945," *Evolution and Human Behaviour* 38. 2 (2017): 208 - 216.

<sup>⑥</sup> Campbell, C. D., & Lee, J. Z. "Price Fluctuations, Family Structure and Mortality in Two Rural Chinese Populations: Household Responses to Economic Stress in Eighteenth- and Nineteenth-century Liaoning," In T. Bengtsson & O. Saito, Eds., *Population and the Economy: From Hunger to Modern Economic Growth* (Oxford: Oxford University Press, 2000), pp. 371 - 420.

<sup>⑦</sup> Campbell, C. D., & Lee, J. Z. "Mortality and Household in Seven Liaodong Populations," In T. Bengtsson, C. D. Campbell, J. Z. Lee et al., *Life under Pressure: Mortality and Living Standards in Europe and Asia, 1700 - 1900* (Cambridge, MA: MIT Press, 2004), pp. 293 - 324.

<sup>⑧</sup> Bengtsson, T., Campbell, C. D., Lee, J. Z. et al., *Life under Pressure: Mortality and Living Standards in Europe and Asia, 1700 - 1900* (Cambridge, MA: MIT Press, 2004).

<sup>⑨</sup> Campbell, C. D., & Lee, J. Z. "Demographic Impacts of Climatic Fluctuations in Northeast China, 1749 - 1900"

初、中、后期不同北京的死亡率，评估了 20 世纪初和 1949 年后北京公共卫生干预措施的效果。<sup>①</sup>

最近的研究考察了家庭背景和家庭历史对老年人及后代死亡率的影响。陈爽使用 CMGPD-SC，比较了双城移民中原籍为北京城市及其周边地区者，和原籍为东北农村者的死亡率差异。<sup>②</sup>她发现，尽管享有国家政策给予的特权，原籍北京的移民后代死亡率仍一直更高。2009 年，康和李使用 CMGPD-LN 研究了家庭环境对成年人和老年人死亡率的影响。他们发现，幼年丧母或出生时母亲年龄在 35 岁以上的男性成年后死亡率更高。那些出生时间与前一胎间隔较短、母亲年龄在 35 岁以上、父亲是残疾人、或父亲是受薪官员的男性，老年时的死亡风险会升高。<sup>③</sup>2014 年，董浩和李中清使用 CMGPD-LN 研究了童年时曾经历迁居的男性的死亡率，发现在迁入地有亲属者迁居后死亡率更低。<sup>④</sup>2018 年，臧晓露和康文林使用 CMGPD-LN 研究了童年时与祖父母共居对成年和老年时死亡率的影响。<sup>⑤</sup>

### 3. 婚姻和家庭

学界对婚姻的研究一直颇为关注，因为婚姻时机和总体成婚几率密切反映了家庭的优先事项和家庭中的个人特权。婚姻直接体现了家庭对子女婚配时间及婚配对象的明确抉择。相比之下，生育率和死亡率（排除杀婴后）虽同样能反映家庭优先事项和决策，但在其他各种因素的影响下，这种关联难以简单解释。团队的研究表明，地位高的男性更有可能结婚，且丧偶后更有可能再婚。1997 年，李中清和康文林首次证明道义男性的社会经济地位、家庭地位与成婚几率之间存在正相关关系。<sup>⑥</sup>在皇室成员中，社会地位也与男性成婚几率正相关。<sup>⑦</sup>共居远亲的社会经济地位同样影响男性成婚几率，此外也有明显证据表明家庭内同一代未婚男性之间存在婚配优先次序。<sup>⑧</sup>地位较高的女性倾向于晚婚，但几乎所有女性最终都会结婚。<sup>⑨</sup>再婚的机会也与社会经济地位有关，地位高的鳏夫更有可能再婚。

9,” In S. Kurosu, T. Bengtsson, & C. D. Campbell, Eds., *Demographic Responses to Economic and Environmental Crisis* (Kashiwa: Reitaku University Press, 2010a), pp. 107 - 132.

<sup>①</sup> Campbell, C. D. “Public Health Efforts in China Before 1949 and Their Effects on Mortality: The Case of Beijing,” *Social Science History* 21. 2 (1997): 179 - 218; Campbell, C. D. “Mortality Change and the Epidemiological Transition in Beijing, 1644 - 1990,” In T.-j. Liu, J. Lee, D. S. Reher, O. Saito, & F. Wang, Eds., *Asian population history* (Oxford: Oxford University Press, 2001), pp. 221 - 247.

<sup>②</sup> Chen, S., Campbell, C. D., & Lee, J. Z. “Vulnerability and Resettlement: Mortality Differences in Northeast China by Place of Origin, 1870 - 1912, Comparing Urban and Rural Migrants,” *Annales de Démographie Historique* 2 (2005, publ. 2006): 47 - 79.

<sup>③</sup> Campbell, C. D., & Lee, J. Z. “Long-term Mortality Consequences of Childhood Family Context in Liaoning, China, 1749 - 1909,” *Social Science and Medicine* 68. 9 (2009): 1641 - 1648.

<sup>④</sup> Dong, H., & Lee, J. Z. “Kinship matters: Long-term Mortality Consequences of Childhood Migration, Historical Evidence From Northeast China, 1792 - 1909,” *Social Science and Medicine* 119 (2014): 274 - 283.

<sup>⑤</sup> Zang, E., & Campbell, C. D. “Males' Later Life Mortality Consequences of Coresidence With Paternal Grandparents: Evidence From Northeast China, 1789 - 1909,” *Demography* 55. 2 (2018): 435 - 457.

<sup>⑥</sup> Lee, J. Z., & Campbell, C. D., *Fate and Fortune in Rural China: Social Organization and Population Behavior in Rural Liaoning, 1774 - 1873* (Cambridge: Cambridge University Press, 1997), pp. 133-156, 177-195.

<sup>⑦</sup> Lee, J. Z., Wang, F., & Ruan, D. “Nuptiality Among the Qing Nobility: 1600 - 1900,” In T.-j. Liu, J. Z. Lee, D. Reher, O. Saito, & F. Wang, eds., *Asian Historical Demography* (Oxford: International Studies in Demography, Oxford University Press, 2001), pp. 353 - 373.

<sup>⑧</sup> Campbell, C. D., & Lee, J. Z. “Kin Networks, Marriage, and Social Mobility in Late Imperial China,” *Social Science History* 32. 2 (2008b): 175 - 214.

<sup>⑨</sup> Chen, S., Campbell, C. D., & Lee, J. Z. “Categorical Inequality and Gender Difference: Marriage and Remarriage in Northeast China, 1749 - 1913,” In C. Lundh, S. Kurosu et al, Eds., *Similarity in difference: Marriage in Europe and Asia, 1700 - 1900* (Cambridge, MA: MIT Press, 2014), pp. 393 - 438.

我们还考虑了婚姻的其他影响因素，包括多妻制和经济冲击。尽管在 20 世纪以前，一夫多妻制是中国最广为人知的婚姻特征之一，但在 CMGPD-LN 和 CMGPD-SC 所涵盖的农村人口中却极为罕见。随着时间推移，即使在皇族中，一夫多妻者也越来越少，因此到 19 世纪后半期，除了皇室宗亲外，一夫多妻者已经非常罕见。此外，一夫多妻制主要用于延长男性的生育时间，而非同时与不同的妻妾生育子嗣。<sup>①</sup>在辽宁农村，粮价上涨带来的经济困难对婚姻的影响存在滞后性，而并不像对死亡率和生育率的影响那样直接——女婴和女童死亡率的上升不均衡地减少了 20 年后成年女性的数量，加剧了婚姻市场的不平衡。<sup>②</sup>

最近，团队研究了婚配对象选择问题，以深入了解家庭对其姻亲的选择偏好。这有助于划定历史上中国各群体之间的社会、经济和制度界限。我们关于该主题的第一篇论文研究了 CMGPD-SC 中的族际婚姻，以了解在制度不禁止八旗子弟与普通平民通婚的情况下，汉族和满族会不会通婚。<sup>③</sup>我们发现，满汉通婚很普遍，其可能性取决于家庭特征，包括家庭异族通婚史、当地婚姻市场构成以及其他因素。我们关于该主题的第二篇论文研究了 20 世纪中期中国山西农村中，教育和家庭阶层因素对婚配选择的影响。<sup>④</sup>研究表明两者均对婚配有重要影响，而且影响机制在 1949 年新中国成立前后变化并不大。这是一个新发现，因为学界虽然此前已有许多关于 20 世纪后半叶中国婚姻双方教育匹配度的研究，但关于 20 世纪中期阶级成分对婚配的作用的研究较少。

另一项工作是调查家庭动态，包括家庭扩张和分家。在辽宁，很大一部分人口生活在宗族中，和众多远亲生活在一起。<sup>⑤</sup>宗族等级森严，成员的地位和特权取决于与族长的关系。<sup>⑥</sup>族长及其子孙最有特权，而较远的亲属则享有较少特权。宗族的分裂，通常由能够联结家族各脉的族长或长老之死所致。<sup>⑦</sup>族长主要是男性，但寡妇有时会在丧夫后继任族长。对族长的远房亲属来说，分家是一种解放：他们以前处于家族底层，分家后却有可能掌控新组成的家庭的资源。<sup>⑧</sup>

## （二）比较

1994 年到 2014 年的二十年间，我们和王丰一起参与了国际合作比较项目——欧亚人口和家庭史项目（Eurasia Project in Population and Family History），并在其中分析了

<sup>①</sup> Lee, J. Z., Wang, F., & Ruan, D. "Nuptiality Among the Qing Nobility: 1600 - 1900," In T.-j. Liu, J. Z. Lee, D. Reher, O. Saito, & F. Wang, eds., *Asian Historical Demography* (Oxford: International Studies in Demography, Oxford University Press, 2001), pp. 353 - 373.

<sup>②</sup> Campbell, C. D., & Lee, J. Z. "Economic Conditions and Male First Marriage in Northeast China, 1749 - 1909," *Sungkyun Journal of East Asian Studies* 8. 1 (2008a): 17 - 42.

<sup>③</sup> Chen, B., Campbell, C. D., & Dong, H. "Interethnic Marriage in Northeast China, 1866 - 1913," *Demographic Research* 38. 34 (2018): 929 - 966.

<sup>④</sup> Xing, L., Campbell, C. D., Li, X., Noellert, M., & Lee, J. Z. "Education, Class and Assortative Marriage in Rural Shanxi, China in the Mid-twentieth Century," *Research in Social Stratification and Mobility* 66 (2020): 1 - 15.

<sup>⑤</sup> Lee, J. Z., & Campbell, C. D. *Fate and Fortune in Rural China: Social Organization and Population Behaviour in Rural Liaoning, 1774 - 1873* (Cambridge: Cambridge University Press, 1997), pp. 105-132; Lee, J. Z., & Gjerde, J. "Comparative Household Morphology of Stem, Joint, and Nuclear Household Systems: Norway, China, and the United States," *Continuity and Change* 1. 1 (1986): 89 - 111.

<sup>⑥</sup> Lee, J. Z., & Campbell, C. D. *Fate and Fortune in Rural China: Social Organization and Population Behaviour in Rural Liaoning, 1774 - 1873* (Cambridge: Cambridge University Press, 1997), pp. 133-156.

<sup>⑦</sup> Lee, J. Z., & Campbell, C. D. "Headship Succession and Household Division in Three Chinese Banner Serf Populations, 1789 - 1909," *Continuity and Change* 13. 1 (1998): 117 - 141.

<sup>⑧</sup> 康文林、李中清：《中国东北地区分家的原因和结果（1789-1909）》，李中清、郭松义、定宜庄主编《婚姻家庭与人口行为》，北京大学出版社，2000，第 1 - 32 页。

---

CMGPD 数据库。通过对欧洲和亚洲的数据集进行几乎相同的分析，该研究比较了瑞典西南部、日本东北部、中国东北部、比利时东部和意大利北部人口对经济状况的反应模式。我们与研究瑞典西南部斯堪尼亚的 Tommy Bengtsson、Christer Lundh 和 Martin Dribe，研究日本东北部福岛的速水融、津谷典子 (Noriko Tsuya) 和黑须里美 (Satomi Kurosu)，研究比利时东部的 Michel Oris 和 George Alter，以及研究意大利北部的 Marco Breschi、Matteo Manfredini 和 Renzo Derosas 展开了广泛交流。<sup>①</sup>

对死亡率、生育率和结婚率的比较研究，揭示出无论在东方还是西方，家庭背景都在塑造人口现象方面发挥着作用。<sup>②</sup>除了这个意料之外的相似点外，东西方之间也存在意料之外的差异。我们发现，影响人口对经济冲击之反应的因素中，在西方，社会经济差异很重要；而在东方，家庭背景的社会政治差异更为重要。总体而言，东方人口对经济冲击的反应比西方弱，这与基于马尔萨斯人口动态解释的预期相反。我们强调使用相同模型分析所有不同数据库以得出比较结果，并由此得出新观点，这使得我们的研究明显区别于以往的国际人口家庭比较研究。

为了深入了解 CMGPD-LN 所用八旗户口登记册的优缺点，我们还比较了 CMGPD-LN 中记录的家庭和他们自己的家谱。在我们的八次实地考察中，从每个村庄收集的材料中都有家族族谱。我们将这些资料转录成一个数据库，然后比较 CMGPD-LN 和家谱的家族成员记录，发现在婴儿期和儿童期死亡的子女往往被排除在家谱之外，从而导致依据家谱估计的生育率低于实际生育水平。<sup>③</sup>我们还发现根据家谱估计的生育率之所以偏低，也因为家谱更有可能忽略那些始终未婚的和婚后无子的成年人。既往研究认为，调整婴儿和儿童死亡率以及出生时的性别比，可以处理忽略女儿及早逝儿子带来的生育率估计值偏误，从而“修正”从家谱估计出的生育率。这项研究则发现，家谱所忽略的无子女成年人带来的偏误，使这种“修正”变得更加困难甚至不可能。

在董浩的带领下，我们新启动了一项关于历史上东亚地区家庭和人口现象的比较性合作研究。董浩整合了来自中国东北、台湾、日本东北和韩国的数据库，并与黑须里美和杨文山 (Wenshan Yang) 合作分析这些数据。在这些比较中，我们还使用了韩国丹城县的家庭户籍簿，这些户籍簿由当时一批主要来自成均馆大学的历史学家公开提供，我们通过姓名连接将其转化为纵向数据。<sup>④</sup>由此比较研究了各东亚人群之间，家庭背景（包括是否拥有特定亲属）对人口行为的影响。<sup>⑤</sup>

---

<sup>①</sup> 每个团队都有一些我们偶尔会进行互动的其他参与者。

<sup>②</sup> 相关研究可参见 Bengtsson, T., Campbell, C. D., Lee, J. Z. et al, *Life under Pressure: Mortality and Living standards in Europe and Asia, 1700 - 1900* (Cambridge, MA: MIT Press, 2004); Tsuya, N.O., Wang, F., Alter, G., Lee, J. Z. et al. *Prudence and Pressure: Reproduction and Human Agency in Europe and Asia, 1700 - 1900* (Cambridge, MA: MIT Press, 2010); Lundh, C., Kurosu, S. et al. *Similarity in difference: Marriage in Europe and Asia, 1700 - 1900* (Cambridge, MA: MIT Press, 2014).

<sup>③</sup> Campbell, C. D., & Lee, J. Z. “State Views and Local Views of Population: Linking and Comparing Genealogies and Household Registers in Liaoning, 1749 - 1909,” *History and Computing* 14(1-2) (2002a, publ. 2006): 9 - 29.

<sup>④</sup> 康文林和董浩编写了程序，将不同登记簿中同一人的记录纵向连接起来，将横截面数据转化为面板数据，参见 Dong, H., Campbell, C. D., Kurosu, S., Yang, W., & Lee, J. Z. “New Sources for Comparative Social Science: Historical Population Panel Data From East Asia,” *Demography* 52. 3 (2015b): 1061 - 1088。建立的纵向连接详见：<https://doi.org/10.14711/dataset/IVIDZV>。

<sup>⑤</sup> Dong, H., Campbell, C. D., Kurosu, S., & Lee, J. Z. “Household Context and Individual Departure: The Case of Escape in Three 'Unfree' East Asian Populations, 1700 - 1900,” *Chinese Journal of Sociology* 1. 4 (2015a)

### (三) 社会流动、不平等和移民

我们对社会流动的研究，始于分析父子成就的关联，进而发展至研究亲属网络对个人成就的影响，最后扩展到研究家族的影响。家族是中国历史上一个关键的社会分层因素。最初对父子成就关系的研究表明，在道义，如果父亲能获得官职，儿子也会有更大的机会获得官职。<sup>①</sup>但与学界对 19 世纪北美和欧洲社会流动的研究结果相比，辽宁地方精英带给儿子的成就优势远没有西方那么明显。<sup>②</sup>民族流动伴随着社会流动，担任官职的汉族男子更有可能将汉族名字改成满族名字。<sup>③</sup>在辽宁的每一代人中，有很大一部分获得政府官职男子是“新”的——他们的父亲没有任何官职，其他父系亲属也没有。<sup>④</sup>拥有其他官员亲属通常也有利于个人官职获得，但并非永远如此。<sup>⑤</sup>

在辽宁农村，家族也是地位分化的一个影响因素。取得成就的机会和结婚的机会，不仅取决于个人和家庭特征，也取决于家族归属。<sup>⑥</sup>家族在当地社会的相对地位，从清代到上世纪末具有长期延续性。<sup>⑦</sup>社会经济特权不仅增加了一个男人的子女数量，而且增加了他最多六代之后的后代总数，这意味着当地每一代人中，前几代最具社会经济特权成员的后代们都在全部人口中占据更高比例。<sup>⑧</sup>我们还探索了计算机技术在家族研究中的应用：傅四维(Siweu Fu)和董浩等使用可视化和网络技术来研究家族繁衍结构的决定因素。<sup>⑨</sup>

最近我们从一个更广泛的角度研究不平等问题。陈爽研究了双城基于体制归属和土地持有的社会分层问题。国家根据体制归属定义的人口类别规定了不同的土地权利。这些不同的土地权利影响了持有土地以及获得其他社会和经济特权的机会。双城居民在某些情况下对国家规定的社会等级制度提出了挑战，但同时也在其他情况下强化了这种制度。<sup>⑩</sup>倪志宏研究了 1945 年后双城土地改革中的个人数据，发现地方强人手中的权力被重新分配，

---

:513 - 539; Dong, H., Campbell, C. D., Kurosu, S., Yang, W., & Lee, J. Z. "New Sources for Comparative Social Science: Historical Population Panel Data From East Asia," *Demography* 52. 3 (2015b):1061-1088; Dong, H. Patriarchy, Family System and Kin Effects on Individual Demographic Behaviour Throughout the Life Course: East Asia, 1678-1945 (Ph. D. diss., Hong Kong University of Science and Technology Division of Social Science, 2016); Dong, H., Manfredini, M., Kurosu, S., Yang, W., & Lee, J. Z. "Kin and Birth Order Effects on Male Child mortality: Three East Asian Populations, 1716 - 1945," *Evolution and Human Behaviour* 38. 2 (2017):208 - 216.

<sup>①</sup> Lee, J. Z., & Campbell, C. D. *Fate and Fortune in Rural China: Social Organization and Population Behaviour in Rural Liaoning, 1774 - 1873* (Cambridge: Cambridge University Press, 1997), pp. 196-215.

<sup>②</sup> Campbell, C. D., & Lee, J. Z. "Social Mobility From a Kinship perspective: Rural Liaoning, 1789 - 1909," *International Review of Social History* 48. 1 (2003): 1 - 26.

<sup>③</sup> Campbell, C. D., Lee, J. Z., & Elliott, M. "Identity Construction and Reconstruction: Naming and Manchu Ethnicity in Northeast China, 1749 - 1909," *Historical Methods* 35. 3 (2002):101 - 116.

<sup>④</sup> Campbell, C. D., & Lee, J. Z. "Social Mobility From a Kinship Perspective: Rural Liaoning, 1789 - 1909," *International Review of Social History* 48.1 (2003): 1 - 26.

<sup>⑤</sup> Campbell, C. D., & Lee, J. Z. "Kin Networks, Marriage, and Social Mobility in Late Imperial China," *Social Science History* 32. 2 (2008b):175 - 214.

<sup>⑥</sup> Campbell, C. D., & Lee, J. Z. "Villages, Descent Groups, Households and Individual Outcomes in Rural Liaoning, 1789 - 1909," In T. Bengtsson & G. Mineau, Eds., *Kinship and Demographic Behaviour in the Past*. (Springer, 2008c), pp. 73 - 104.

<sup>⑦</sup> Campbell, C. D., & Lee, J. Z. "Kinship and the Long-term Persistence of Inequality in Liaoning, China, 1749 - 2005," *Chinese Sociological Review* 44. 1 (2011): 71 - 104.

<sup>⑧</sup> Song, X., Campbell, C. D., & Lee, J. Z. "Ancestry Matters: Patrilineage Growth and Extinction," *American Sociological Review* 80.3 (2015):574 - 602.

<sup>⑨</sup> Fu, S., Dong, H., Cui, W., Zhao, J., & Qu, H. "How Do Ancestral Traits Shape Family Trees Over Generations?" *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018): 205 - 214.

<sup>⑩</sup> Chen, S. *State-Sponsored Inequality: The Banner System and Social Stratification in Northeast China* (Stanford: Stanford University Press, 2017)



---

这为财产的重新分配铺平了道路——新分配仍然由国家界定。<sup>①</sup>

团队还研究了移民问题。CMGPD-LN 跟踪了在辽宁省内迁移的家庭。离开原居住地的行为通常是非法的，但也会被记录下来。我们的第一项相关研究考察了影响辽宁省内家庭合法迁移和非法迁出该地区的因素。<sup>②</sup>家庭年龄结构制约着合法移民，老人较少的“年轻”家庭更有可能移民。同时，官员家庭不太可能迁移。非法迁移在未婚或丧偶的男性、族长的远房亲戚或小家庭的成员中更为常见。董浩等人比较研究了中国东北地区的移民模式与 18 和 19 世纪的韩国、日本的移民模式。<sup>③</sup>

#### （四）20 世纪中国大学生的社会与地理来源

通过对 CUSD 学籍卡数据库等其他材料的研究，团队阐释了从 19 世纪末到 21 世纪初中国大学生的地理来源和社会出身的变化。在清代，1905 年废除科举考试之前，教育精英经由科举制度在全国范围内遴选。而在 20 世纪上半叶的中华民国时期，教育精英大多来自沿海大城市的商人和专业人士家庭。<sup>④</sup>此外，梁晨等人发现，新中国成立初期，北京大学和苏州大学的学生来源仍然类似于民国时期，来自沿海城市的商人和专业人才家庭的学生比例高。<sup>⑤</sup>

更重要的是，梁晨等还表明，1955 年引入的标准化考试（高考），以及中小学教育的大幅扩张，从根本上改变了大学生源构成。农民和工人家庭首个大学生的数量在这一时期显著增加，这种情况一直持续到 20 世纪 90 年代。上世纪 90 年代，专业人士家庭出身的大学生比例开始回升。然而，至少到 2004 年，北京大学约 30% 的学生和苏州大学约 40% 的学生仍然来自工人家庭。<sup>⑥</sup>这与西方大部分国家的模式截然不同：在西方，就读于精英私立大学的学生与北大、苏大学生不同，绝大多数来自高收入家庭。高考是为来自普通家庭的学生保留了机会，还是更偏向于来自富裕家庭的学生？这是当下的热门话题，这些发现为该辩论提供了重要论据。<sup>⑦</sup>

#### （五）清代官场与宦宦生涯

通过分析 CGED-Q，我们深入了解了清代官场和官员生涯。这是研究清代官员的传统

---

<sup>①</sup> Noellert, M. *Power over Property: The Political Economy of Communist Land Reform in China* (Ann Arbor: University of Michigan Press, 2020).

<sup>②</sup> Campbell, C. D., & Lee, J. Z. "Free and Unfree Labour in Qing China: Emigration and Escape Among the Bannermen of Northeast China, 1789 - 1909," *The History of the Family* 6. 4 (2001): 455 - 476.

<sup>③</sup> Dong, H., Campbell, C. D., Kurosu, S., & Lee, J. Z. "Household Context and Individual Departure: The Case of Escape in Three 'Unfree' East Asian Populations, 1700 - 1900," *Chinese Journal of Sociology* 1.4 (2015a): 513 - 539.

<sup>④</sup> 梁晨、董浩、任韵竹、李中清：《江山代有才人出——中国教育精英的来源与转变（1865—2014）》，《社会学研究》2017 年第 3 期；Bamboo Y. Ren, Chen Liang, James Z. Lee, "Meritocracy and the Making of the Chinese Academe, 1912-1952," *China Quarterly* 244 (2020): 942-968.

<sup>⑤</sup> 梁晨、张浩、李兰等著：《无声的革命 北京大学、苏州大学学生社会来源研究 1949-2002》，生活·读书·新知三联书店，2013。

<sup>⑥</sup> 具体请参考梁晨、李中清、张浩等：《无声的革命：北京大学与苏州大学学生社会来源研究（1952-2002）》，《中国社会科学》2012 年第 1 期。梁晨、张浩、李兰等著：《无声的革命 北京大学、苏州大学学生社会来源研究 1949-2002》，生活·读书·新知三联书店，2013。

<sup>⑦</sup> 2012 年发表的一篇公开这些结果的文章（即发表在《中国社会科学》上的《无声的革命》）立即引起了公众的广泛讨论，我们在团队网页上对这些讨论进行了总结。共计有超过 100 篇相关文章在 1000 多个中国媒体网站上被转发、转载或转播，其中包括一些电视报道和讨论。最引人关注的是，全国政协主席俞正声在 2014 年中国人民政治协商会议的一场教育分会提到了我们的书《无声的革命》。2020 年 7 月 12 日，余亮教授在中国的 Bilibili 网站上发布了关于该书的视频评论，截至 2020 年 7 月 15 日该视频已经被浏览了 24 万次，被点赞了 2.2 万次。

---

方法所不及的，因为传统方法强调对个人、职位或特定时间段的案例研究。任玉雪、陈必佳、康文林、李中清等人的研究表明，中央政府，尤其是上层，直到清末都由满人和其他旗人主导。只有一小部分科举出身的汉人得以在中央政府任职，且主要局限于翰林院及其相关部门。然而在中央政府之外，官员主要是汉人，且有更多官员出身于捐纳而非科举。<sup>①</sup>所有官员仕途长短的中位数不到七年，其中封疆大吏和贡生仕途生涯的中位数仅为三年。<sup>②</sup>1905年科举制度的废除，对已获科名者的仕途没有什么影响。陈必佳、康文林、李中清研究了清代末年的旗人官员，发现他们的人数和职位在清末新政时期变化不大，但由于民人官员人数的增加，他们在官员中的比例下降了。<sup>③</sup>2020年，康文林的研究表明，在1905年废除科举考试后，已获科名候缺待任者以跟从前相同的速度进入仕途，而已获官职者的流动率则没有受到影响。<sup>④</sup>这一发现挑战了学界科举停废阻碍精英抱负施展的观点。<sup>⑤</sup>

## 五 总结

回顾四十年来在人口、社会和经济史方面的协作研究，以下是一些反思和见解。首先，在寻找、获取和构建多样化大型微观数据集方面，我们是极幸运的，这几乎是我们所有研究的基础。这是团队努力的结果。在搜寻微观数据以理解古今中国方面，我们所取得的成就日益依赖于与李-康研究团队各位同事的合作。典型的例子包括：梁晨和任韵竹对民国时期大学生学籍卡的整理编纂，任玉雪对她所做缙绅录研究进展的分享，倪志宏和李湘宁对20世纪中期中国乡村建设相关新材料的发掘，任韵竹、吴艺贝和杨莉关于海外留学生和专业人士材料的最新发现等。

其次，诸多机构的支持，对我们获得本文所讨论的微观数据至关重要。辽宁省档案馆、第一历史档案馆、辽宁省地方志办公室、犹他州家谱协会（现称为 FamilySearch）、山西大学中国社会历史研究中心、双城县档案馆等诸多机构，以及中国和美国的许多大学，如北京大学、苏州大学、上海交通大学、清华大学、浙江大学和哥伦比亚大学等，均向我们慷慨提供了馆藏资料。另有一些数据库，尤其是 CGED-Q，依靠于已经出版或公开提供下载渠道的材料。其中最重要的是清华大学图书馆馆藏清代《缙绅录》，它与哈佛大学燕京图书馆和哥伦比亚大学图书馆馆藏《缙绅录》一起，构成了 CGED-Q 的主要数据来源。

第三，许多学界同仁对团队这些数据库的支持，也使我们深深受益。囿于篇幅，无法列出他们中的所有，在此谨单独列出一些发挥了关键作用的人物。鞠德源、Robert Eng、Alice Suen、迟少艾(Anna Chi)等人帮助李中清启动了对道义的研究。Mel Thatcher 安排查

---

<sup>①</sup> 相关研究可参见，任玉雪、陈必佳、郝小雯、康文林、李中清：《清代缙绅录量化数据库与官员群体研究》，《清史研究》2016年第4期；Chen, B. *Origins and Career Patterns of the Qing Government Officials (1850–1912): Evidence from the China Government Employee Dataset-Qing (CGED-Q)* (Ph. D. diss., Hong Kong University of Science and Technology Division of Social Science, 2019)；Chen, B., Campbell, C. D., Ren, Y., & Lee, J. Z. “Big Data for the Study of Qing Officialdom: The China Government Employee Database-Qing (CGED-Q),” *The Journal of Chinese History* Volume 4, Special Issue 2(2020):431 - 460.

<sup>②</sup> Chen, B., Campbell, C. D., Ren, Y., & Lee, J. Z. “Big Data for the Study of Qing Officialdom: The China Government Employee Database-Qing (CGED-Q),” *The Journal of Chinese History* Volume 4, Special Issue 2(2020):431 - 460.

<sup>③</sup> 陈必佳、康文林、李中清：《清末新政前后旗人与宗室官员的官职变化初探——以〈缙绅录〉数据库为材料的分析》，《清史研究》2018年第4期。

<sup>④</sup> 康文林：《清末科举停废对士人文官群体的影响——基于微观大数据的宏观新视角》，《社会科学辑刊》2020年第4期。

<sup>⑤</sup> Bai, Y., & Jia, R. “Elite Recruitment and Political Stability: The Impact of the Abolition of China's Civil Service Exam System,” *Econometrica* 84. 2 (2016): 677 - 733.

---

阅了犹他州家谱协会馆藏。台湾“中研院”的刘翠溶 (Ts'ui-jung Liu)、刘素芬和赖惠敏协助建立了 CMGPD-IL 和 CMGPD-LN。在“中研院”的录入人员中, 蔡淑美(Shu-mei Tsay)对 CMGPD-IL 和 CMGPD-LN 的贡献最大。陈爽、倪志宏和陈必佳分别协调和监督了 CMGPD-SC、CRRD-LR 和 CRRD-SQ 以及 CGED-Q 的数据输入工作。梁晨、任韵竹、张浩、吴艺贝和杨莉发起、协调并监督了 CUSD 和 CPOD 的各个子数据库的建立。董浩帮助创建了韩国登记册的纵向连接, 并领导协 CMGPD 和日本、韩国及台湾的其他数据库的整合工作。大量录入人员孜孜不倦地输入所有这些数据。这里无法列出所有录入人员, 但须得特别强调六位长期做出极大贡献的人。孙惠成、激扬和肖星输入了大量的 CMGPD-LN、CMGPD-SC、CRRD-LR 数据, 并与葛晓东、刘北箴和赵宓一起输入了 CGED-Q 的数据。

第四, 如果没有慷慨的机构支持和零星的个人支持, 我们不可能坚持至今。李中清在加州理工学院开始了他的职业生涯, 正是在那里, 当时还是本科生的康文林与他结识。现在回想起来, 加州理工学院是仅有的几个能在 20 世纪 80 年代初支持一位人文/历史学助理教授(后来升任正教授)进行中国问题定量研究的地方之一。也很难想象, 在任何其他学术机构, 一个学习电子工程、没有中文语言能力的大二学生, 仅仅出于对中国历史的兴趣, 就可以走进历史学教授的办公室、经过一番讨论、勾勒一个计划、为一个正在进行的项目重新组织数据管理和分析流程——并从此成为一位合作者。康文林随后在宾夕法尼亚大学研究生院学习, 然后在加州大学洛杉矶分校担任社会学助理教授、副教授和正教授, 尽管他的工作内容不那么容易理解, 他仍获得了导师和同事的认可。

来自加州理工学院、密歇根大学、加州大学洛杉矶分校、北京大学、香港科技大学、上海交通大学以及中国其他大学的内部资金和行政支持, 以及美国国家卫生研究院、中国大陆国家自然科学基金、香港研究资助局以及台湾科学委员会的持续研究支持, 使团队的数据采集和数据库建设能力大大提升。同样重要的是, 各大学为团队提供了与研究生、博士后和客座教授合作的机会(有时还提供了资金), 他们为过去四十年间的数据库建设和研究做出重要贡献, 并在一些项目中发挥了领导作用, 希望未来仍能如此。特别感谢 Myron Guttman, 作为校际政治和社会研究联盟(ICPSR)的负责人, 他在我们扩展工作规模及首次寻求大量外部资金时提供了指导和支持。

长期合作对我们的工作起到了关键推动作用。对我们来说, 最持久、最有影响力的合作, 是我们与来自不同国家和学科的同事在“人口和家庭史的欧亚项目”上共事了 20 年。与其他项目参与者的互动, 促使我们扩大研究课题范围、学习和应用更先进的方法、以及为其他项目寻找开展比较研究的机会。与其他研究同类数据或课题的同人们在频繁、持续互动中产生的友情, 鼓舞了团队成员的士气。我们对与王丰长达 20 年、卓有成效的合作记忆犹新, 这种合作零散地产生了许多研究成果, 比如《人类的四分之一》(One Quarter of Humanity)及 *Prudence and Pressure* 等书。<sup>①</sup>我们与 Tommy Bengtsson 和津谷典子(Noriko Tsuya)也有合作, 包括我们在隆德和东京以及他们在帕萨迪纳的相互来访。

---

<sup>①</sup> Lee, J. Z., & Wang, F., *One Quarter of Humanity: Malthusian Mythology and Chinese Realities, 1700 - 2000* (Cambridge, Mass.: Harvard University Press, 1999); Tsuya, N.O., Wang, F., Alter, G., Lee, J. Z. et al, *Prudence and pressure: Reproduction and Human Agency in Europe and Asia, 1700 - 1900.* (Cambridge, MA: MIT Press, 2010).

---

许多关于特定论文和长期项目上的短期合作也同样重要。定宜庄和郭松义针对 CMGPD-LN 提供了建议，在辽宁省地名办公室高静的帮助下，我们与他们一起进行了实地考察，最终形成了一本书。<sup>①</sup>在我们构建和分析 CMGPD-IL 时，郭松义也与我们分享了他的专业知识。我们也有幸与其他许多人共同撰写了使用我们数据库的论文或论文集，这些合作者包括 Lawrence Anthony、欧立德 (Mark Elliott)、Robert Eng、William Lavelly、马文清 (Chris Myers)、宋曦 (Song Xi)、陈慧雯 (Alice Suen)、谭国富 (Guofu Tan)、臧晓露 (Emma Zang) 和傅四维 (Siwei Fu)，以及屈华民 (Huamin Qu) 团队的其他成员。同样，我们也从与速水融、金建泰 (Kuen-tae Kim)、黑须里美、朴铉濬 (Hyunjoon Park)、李相国 (Sangkuk Lee)、刘翠溶、孙炳圭 (Byun-giu Son)、津谷典子、杨文山及其他合作者的交流互动中受益。

回顾过去，我们认为，团队所做研究的一个显著特点，即使用数据归纳和数据驱动方法，对我们取得成功至关重要。这种方法强调通过对我们构建的数据库进行实证分析，发现有关人口现象、家庭、社会和经济组织的规律。我们总是从寻找有助于研究学界普遍感兴趣的某个话题的数据开始。接下来，通过探索性和描述性分析，我们试图揭示人口现象、家庭及社会组织的关键模式。只有在开展大量工作来验证数据，进而详细阐释描述性分析中发现的关系和模式之后，我们才会转向精心设计的回归模型。虽然这种方法很耗时，有时需要花费数年来寻找、获取、输入和清理数据，在此基础上进行分析，然后才能取得主要成果，但我们相信，这样的做法使得我们对中国的家庭、社会和经济组织基本模式的理解发生了根本性的转变，而且直到最近这种转变还在不断增加。

相较于解释性学术研究，我们坚信微观数据驱动的实证研究有其重要性，因为关于中国历史仍有许多未知、错知。<sup>②</sup>因此，在力所能及之处，我们进行了尽可能完整的数据录入，投入大量时间和精力来制作详细的文档和用户指南以备数据公开发布，创建了一个完整、永久的资源以供我们自己和其他人研究一系列广泛主题。本文所述的这些项目，正被推广至针对世界上其它地区的研究中。我们期待在未来，构建并应用此类数据库能够成为社会科学和历史学研究的常态——在纵贯式调查和其它统计资料出现之前，这些数据能用来发现历史事实。

[康文林，香港科技大学社会科学部讲座教授、华中师范大学特聘教授；李中清，香港科技大学人文学部及社会科学部讲座教授暨言爱基金社会科学教授]

---

<sup>①</sup> 定宜庄、郭松义、李中清、康文林：《辽东移民中的旗人社会 历史文献、人口统计与田野调查》，上海社会科学院出版社，2004。

<sup>②</sup> 世界范围内，一个很好的例子是马尔萨斯关于中国人口模式的论述，这一理论一度被广泛接受，可参见 Lee, J. Z., & Wang, F., *One Quarter of Humanity: Malthusian Mythology and Chinese Realities, 1700 - 2000* (Cambridge, Mass.: Harvard University Press, 1999)。中国范围内，一个典型案例是关于来自贫困家庭的大学生是否有机会进入一些最具竞争力的大学的辩论。此外，我们还反思了对死亡率和相关行为的误解，并描述了纠正错误及建立新认知的过程，可参见 Bengtsson, T., Campbell, C. D., Lee, J. Z. et al, *Life under Pressure: Mortality and Living Standards in Europe and Asia, 1700 - 1900* (Cambridge, MA: MIT Press, 2004), p.435-439。