**ICS** 11.020

|  |
| --- |
| **C** 07 |
|  |

团体标准

**T/BRA-CDCHE// XXXX-XXXX**

|  |
| --- |
|  |

医学队列平台数据技术

队列数据管理 数据采集质控要求

Medical Cohort Platform Data Technology——

Cohort Data Managements——Requirements of Quality Control on Data Collection

|  |
| --- |
| (草案) |
|  |

20XX- XX - XX发布

20XX - XX - XX实施

北京慢性病防治与健康教育研究会 发布

 目  次

[前  言 II](#_Toc168843443)

[1. 范围 1](#_Toc168843444)

[2. 规范性引用文件 1](#_Toc168843445)

[3. 术语和定义 1](#_Toc168843446)

[4. 采集内容及方式 2](#_Toc168843453)

[5. 质控要求 2](#_Toc168843459)

[5.1 数据采集前 2](#_Toc168843460)

[5.1.1 标准操作程序 2](#_Toc168843461)

[5.1.2 病例报告表 2](#_Toc168843462)

[5.1.3 人员培训 2](#_Toc168843462)

[5.1.4 现场检查 2](#_Toc168843463)

[5.2 数据采集中 2](#_Toc168843464)

[5.2.1 核实调查对象情况 2](#_Toc168843465)

[5.2.2 采集信息 2](#_Toc168843466)

[5.2.3 及时检查数据质量并进行反馈和整改 3](#_Toc168843467)

[5.2.4 确保调查的真实性 3](#_Toc168843468)

[5.2.5 定期评估反馈并作调整 3](#_Toc168843469)

[5.3 数据采集后 3](#_Toc168843470)

[5.3.1 数据清理 3](#_Toc168843471)

[5.3.2 锁定数据库并撰写数据库编码文件 3](#_Toc168843472)

[5.3.3 文件归档 3](#_Toc168843473)

[5.3.4 生物样本的保存、运输和使用 3](#_Toc168843474)

[参考文献 4](#_Toc168843475)

前  言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由北京大学第六医院提出。

本文件由北京慢性病防治与健康教育研究会归口。

本文件起草单位：北京大学第六医院、天津市安定医院、北京大学、中国医学科学院肿瘤医院、山东大学齐鲁医院、北京大学第一医院、中国电子技术标准化研究院、中国疾病预防控制研究中心。

本文件主要起草人：刘肇瑞、徐广明、尹慧芳、黄雨、黄悦勤

本文件其它起草人：邓咏妍、李明慧、魏文强、孙可欣、吕明、李航、李瑞琪、陈园生、杨孝荣、陈浩、白倩倩、王悦。

队列数据管理 数据采集质控要求

1. 范围

本文件规定了在开展跨队列研究前，原始队列数据的采集内容以及质控要求等内容。

本文件适用于指导拟开展跨队列研究的原始队列数据采集工作，包括但不限于社区人群队列、区域性人群队列、针对某一疾病种类或基于特殊机构建立的人群队列。

1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 39767-2021 人类生物样本管理规范

WS/T 306 卫生信息数据集元数据规范

WS 363 卫生信息数据元目录（系列标准）

WS 364 卫生信息数据元值域代码（系列标准）

WS 365 城乡居民健康档案基本数据集

WS/T 370 卫生信息基本数据集编制规范

WS 372-2012 疾病管理基本数据集

WS 375 疾病控制基本数据集（系列标准）

1. 术语和定义

下列术语和定义适用于本文件。

* 1.

队列 cohort

根据某个或某些共同特征而组建的一组特定人群。

注：特征包括但不限于：暴露因素、疾病或健康状态、出生时间或年代、地域、干预措施。

* 1.

跨队列 cross-cohort

队列（3.1）间进行特征数据比较、融合和分析。

注：跨队列形式包括：a）横向跨队列：在不同元数据的队列间进行比较、融合和分析；b）纵向跨队列：在相同元数据的队列间进行比较、融合和分析。

* 1.

调查对象 participants

队列（3.1）中的人，可以是一般人群，也可以是患者。

* 1.

病例报告表 case report form

指按照研究方案要求设计的记录研究对象相关信息的纸质或者电子文件。

* 1.

标准操作程序 standard operating procedures

指为保证队列数据采集工作的一致性而制定的详细的书面要求。

1. 采集内容及方式
	1. 根据研究的目标和需求，明确队列数据采集内容及采集方式。
	2. 采集内容通常包括采集调查对象的社会人口学数据、疾病或健康数据、医疗和照护信息，遗传学信息以及与疾病和健康有关的生活方式和行为等影响因素信息。
	3. 采集方式通常包括自评问卷、他评问卷、体格检查、影像学检查以及生物样本检测。
	4. 应为每一个调查对象确定唯一标识，如身份证号、社会保障号、样本编号或病历号等，队列数据库应清晰易懂，避免信息缺失和重复,以便于后续根据跨队列研究的需要进行数据清理、整合和分析。
	5. 数据应符合卫生健康相关数据集的标准WS/T 306、WS 363、WS 364、WS 365、WS/T 370、WS 372.3、WS 375。
2. 质控要求
	1. 数据采集前
		1. 标准操作程序

数据采集前应设置标准操作程序，明确数据采集的流程、确定数据质控评估标准并建立质量核查方式。

* + 1. 病例报告表

应当确定病例报告表的内容，开发完成数据采集或录入系统，并对数据录入格式及有效值范围进行设定，尽量使用电子化病例报告表，一方面可以采集并行数据以进行质量核查，另一方面可减少录入错误以提高数据质量。

* + 1. 人员培训

数据采集前应对所有参与数据采集和质控的人员进行基本职责培训和专业技能培训，经考核合格后方可参与数据采集或质控工作。数据采集和质控工作应由不同的人员开展。所有数据采集和质控人员还应参加定期培训和阶段性考核，确保所有人员长期保持良好的开展数据采集工作能力，从而保证队列建设高质量地推进。同时，可根据研究要求，对参与人员的资质进行评估，内容包括但不限于其是否具有相关医学教育背景、职称和工作经历等。

* + 1. 现场检查

评估调查现场是否准备妥当，是否符合调查的标准，并对现场评估情况进行记录，现场评估可参照以下内容进行：

* 1. 现场是否整洁，秩序良好；
	2. 项目标识和引导牌是否明确；
	3. 采集人员是否佩戴工作胸卡或身穿统一服装；
	4. 调查所用工具是否准备齐全且完好无损可使用
	5. 实验台面、耗材摆放是否干净且整齐；
	6. 如使用电子化数据采集，还应检查采集设备及采集程序是否可以正常使用。
	7. 数据采集中
		1. 核实调查对象情况

核查调查对象是否符合纳入标准且不符合排除标准，是否具有采集信息的条件，以及调查对象是否签署知情同意等。

* + 1. 采集信息

由数据采集人员按照标准操作程序的数据采集流程采集信息，注意保护调查对象的隐私，主观访谈时不诱导或代替调查对象回答问题，客观检查时应注意保存检查的原始结果。

* + 1. 及时检查数据质量并进行反馈和整改

由质控人员按照标准操作程序的数据质控评估标准以及质量核查方式，及时检查数据质量。尽量在调查对象数据采集后一周内完成数据质量检查。如发现存在质量问题，应及时反馈给数据采集人员，并监督其完成整改。对存在严重质量问题的数据，应重新采集或对该条数据进行废弃标记。对数据进行核查的内容包括但不限于：

1. 数据唯一性：通过比较每名调查对象收集的可识别数据项（如姓名、身份证号、社会保障号）进行核查，可核查数据集内或数据集间不同研究对象的个体唯一性标识和有效记录是否重复。
2. 数据完整性：核查数据集的实际样本量或记录数与应获得数量是否相同；核查变量的完整性，除外重复变量，核实数据集中已有变量数与应获得变量数是否相同；变量值的完整性，应核查数据集中特定单元格信息是否完整，是否有缺失的数据。
3. 数据合法性：对单一数据元进行核查，是否有非法编码，是否超出数据元属性。
4. 内部一致性：对多个数据元之间的逻辑进行核查，是否存在逻辑不合理性（如确诊时间晚于调查对象的年龄等）。
	* 1. 确保调查的真实性

质控人员可以通过等多种方式对调查的真实性进行判断。核查的方式包括但不限于：

1. 数据核查：对数据采集的时长和速度信息进行监控，如时长过短速度过快则有可能调查不真实。
2. 电话核查：致电调查对象，了解调查时的相关情况，并对关键人口学信息进行比对，如调查对象否认接受调查，或关键信息比对失败则有可能调查不真实。
3. 录音核查：在获得调查对象允许的前提下对调查过程进行录音，如录音率过低或录音中声音与调查对象年龄或性别不符则有可能调查不真实。
4. 实地核查：选取部分调查对象，再次进行调查或进行访谈以了解调查过程，如调查主要结果不符或调查对象否认参与调查则有可能调查不真实。
	* 1. 定期评估反馈并作调整

定期生成执行进度报告和质控报告，由项目负责人团队对现场采集的数据进行描述性统计和分析，发现现场执行中的问题（如执行率低、失访率高等），查找系统误差产生的原因，向数据采集人员和质控人员反馈，传达调整建议并督促落实。

* 1. 数据采集后
		1. 数据清理

数据清理时不可在原始数据库中直接操作，应通过撰写数据清理命令完成数据清理工作，并另存为清理后的数据库。数据清理命令中应适当增加注释性文字以便进行后期核查。

* + 1. 锁定数据库并撰写数据元专用属性表

数据清理后应锁定数据库，以便进行跨队列研究。应当撰写数据元专用属性表，即数据字典，内容包括但不限于数据元名称、定义、数据元值的数据类型、表示格式以及数据元允许值。

* + 1. 文件归档

应及时将知情同意书、调查问卷、CRF、临床检验或检查报告单等原始资料归档保存，纸质版原始材料应由专人加密保管，档案保存时限不少于10年。在确保安全的前提下，可以实行电子归档。

* + 1. 生物样本的保存、运输和使用

如采集了生物学样本，生物学样本的保存、运输和使用应按照GB/T 39767执行。

参 考 文 献

1. At, J., R. Bryce, M. Prina, D. Acosta, C. P. Ferri, M. Guerra, Y. Huang, J. J. Rodriguez, A. Salas, A. L. Sosa, J. D. Williams, M. E. Dewey, I. Acosta, Z. Liu, J. Beard and M. Prince (2015). "Frailty and the prediction of dependence and mortality in low- and middle-income countries: a 10/66 population-based cohort study." BMC Med 13: 138.
2. Bernardi, F. A., D. Alves, N. Crepaldi, D. B. Yamada, V. C. Lima and R. Rijo (2023). "Data Quality in Health Research: Integrative Literature Review." J Med Internet Res 25: e41446.
3. Bycroft, C., C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly and J. Marchini (2018). "The UK Biobank resource with deep phenotyping and genomic data." Nature 562(7726): 203-209.
4. Chang, C. K., R. D. Hayes, M. T. Broadbent, M. Hotopf, E. Davies, H. Møller and R. Stewart (2014). "A cohort study on mental disorders, stage of cancer at diagnosis and subsequent survival." BMJ Open 4(1): e004295.
5. Cheikh Ismail, L., H. E. Knight, E. O. Ohuma, L. Hoch and W. C. Chumlea (2013). "Anthropometric standardisation and quality control protocols for the construction of new, international, fetal and newborn growth standards: the INTERGROWTH-21st Project." Bjog 120 Suppl 2(0 2): 48-55, v.
6. Coffey, S., A. J. Lewandowski, S. Garratt, R. Meijer, S. Lynum, R. Bedi, J. Paterson, M. Yaqub, J. A. Noble, S. Neubauer, S. E. Petersen, N. Allen, C. Sudlow, R. Collins, P. M. Matthews and P. Leeson (2017). "Protocol and quality assurance for carotid imaging in 100,000 participants of UK Biobank: development and assessment." Eur J Prev Cardiol 24(17): 1799-1806.
7. de Lusignan, S., S. T. Liaw, P. Krause, V. Curcin, M. T. Vicente, G. Michalakidis, L. Agreus, P. Leysen, N. Shaw and K. Mendis (2011). "Key concepts to assess the readiness of data for international research: data quality, lineage and provenance, extraction and processing errors, traceability, and curation. Contribution of the IMIA Primary Health Care Informatics Working Group." Yearb Med Inform 6: 112-120.
8. Downey, P. and T. C. Peakman (2008). "Design and implementation of a high-throughput biological sample processing facility using modern manufacturing principles." Int J Epidemiol 37 Suppl 1: i46-50.
9. Duda, S. N., B. E. Shepherd, C. S. Gadd, D. R. Masys and C. C. McGowan (2012). "Measuring the quality of observational study data in an international HIV research network." PLoS One 7(4): e33908.
10. Fan, P., S. Zhang, W. Wang, Z. Yang, W. Tan, S. Li, C. Zhu, D. Hu, X. Zhou, Z. Tian, Y. Wang, F. Liu, W. Huang and L. Chen (2022). "The design and implementation of natural population cohort study Biobank: A multiple-center project cooperation with medical consortia in Southwest China." Front Public Health 10: 996169.
11. Gassman, J. J., W. W. Owen, T. E. Kuntz, J. P. Martin and W. P. Amoroso (1995). "Data quality assurance, monitoring, and reporting." Control Clin Trials 16(2 Suppl): 104s-136s.
12. Hilner, J. E., A. McDonald, L. Van Horn, C. Bragg, B. Caan, M. L. Slattery, R. Birch, C. G. Smoak and J. Wittes (1992). "Quality control of dietary data collection in the CARDIA study." Control Clin Trials 13(2): 156-169.
13. Huang, Y., Z. Liu, H. Wang, X. Guan, H. Chen, C. Ma, Q. Li, J. Yan, Y. Yu, C. Kou, X. Xu, J. Lu, Z. Wang, L. Liu, Y. Xu, Y. He, T. Li, W. Guo, H. Tian, G. Xu, X. Xu, S. Lv, L. Wang, L. Wang, Y. Yan, B. Wang, S. Xiao, L. Zhou, L. Li and L. Tan (2016). "The China Mental Health Survey (CMHS): I. background, aims and measures." Soc Psychiatry Psychiatr Epidemiol 51(11): 1559-1569.
14. Huang, Y., Y. Wang, H. Wang, Z. Liu, X. Yu, J. Yan, Y. Yu, C. Kou, X. Xu, J. Lu, Z. Wang, S. He, Y. Xu, Y. He, T. Li, W. Guo, H. Tian, G. Xu, X. Xu, Y. Ma, L. Wang, L. Wang, Y. Yan, B. Wang, S. Xiao, L. Zhou, L. Li, L. Tan, T. Zhang, C. Ma, Q. Li, H. Ding, H. Geng, F. Jia, J. Shi, S. Wang, N. Zhang, X. Du, X. Du and Y. Wu (2019). "Prevalence of mental disorders in China: a cross-sectional epidemiological study." Lancet Psychiatry 6(3): 211-224.
15. Kässens, J. C., L. Wienbrandt and D. Ellinghaus (2021). "BIGwas: Single-command quality control and association testing for multi-cohort and biobank-scale GWAS/PheWAS data." Gigascience 10(6).
16. Kim, S. M., Y. Choi, B. Y. Choi, M. Kim, S. I. Kim, J. Y. Choi, S. W. Kim, J. Y. Song, Y. J. Kim, M. K. Kee, M. Yoo, J. G. Lee and B. Y. Park (2020). "Prospective cohort data quality assurance and quality control strategy and method: Korea HIV/AIDS Cohort Study." Epidemiol Health 42: e2020063.
17. Kühn, A., A. Nieters, A. Köttgen, O. N. Goek, K. Michels, U. Nöthlings, G. Jacobs, C. Meisinger, F. Pessler, M. F. Akmatov, J. Kühnisch, S. Moebus, E. Glocker, S. Naus, M. Keimling, M. Leitzmann, J. Linseisen, H. Sarioglu, C. von Toerne, S. M. Hauck, H. Wallaschofski, H. E. Wichmann and T. Illig (2014). "Feasibility and quality development of biomaterials in the pretest studies of the German National Cohort." Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 57(11): 1255-1263.
18. Kvale, M. N., S. Hesselson, T. J. Hoffmann, Y. Cao, D. Chan, S. Connell, L. A. Croen, B. P. Dispensa, J. Eshragh, A. Finn, J. Gollub, C. Iribarren, E. Jorgenson, L. H. Kushi, R. Lao, Y. Lu, D. Ludwig, G. K. Mathauda, W. B. McGuire, G. Mei, S. Miles, M. Mittman, M. Patil, C. P. Quesenberry, Jr., D. Ranatunga, S. Rowell, M. Sadler, L. C. Sakoda, M. Shapero, L. Shen, T. Shenoy, D. Smethurst, C. P. Somkin, S. K. Van Den Eeden, L. Walter, E. Wan, T. Webster, R. A. Whitmer, S. Wong, C. Zau, Y. Zhan, C. Schaefer, P. Y. Kwok and N. Risch (2015). "Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort." Genetics 200(4): 1051-1060.
19. Leening, M. J., M. Kavousi, J. Heeringa, F. J. van Rooij, J. Verkroost-van Heemst, J. W. Deckers, F. U. Mattace-Raso, G. Ziere, A. Hofman, B. H. Stricker and J. C. Witteman (2012). "Methods of data collection and definitions of cardiac outcomes in the Rotterdam Study." Eur J Epidemiol 27(3): 173-185.
20. Li, M., Y. Huang, Z. Liu, R. Shen, H. Chen, C. Ma, T. Zhang, S. Li and M. Prince (2020). "The association between frailty and incidence of dementia in Beijing: findings from 10/66 dementia research group population-based cohort study." BMC Geriatr 20(1): 138.
21. Liu, Z., Y. Huang, P. Lv, T. Zhang, H. Wang, Q. Li, J. Yan, Y. Yu, C. Kou, X. Xu, J. Lu, Z. Wang, H. Qiu, Y. Xu, Y. He, T. Li, W. Guo, H. Tian, G. Xu, X. Xu, Y. Ma, L. Wang, L. Wang, Y. Yan, B. Wang, S. Xiao, L. Zhou, L. Li, L. Tan, H. Chen and C. Ma (2016). "The China Mental Health Survey: II. Design and field procedures." Soc Psychiatry Psychiatr Epidemiol 51(11): 1547-1557.
22. Llibre Rodriguez, J. J., A. M. Prina, D. Acosta, M. Guerra, Y. Huang, K. S. Jacob, I. Z. Jimenez-Velasquez, A. Salas, A. L. Sosa, J. D. Williams, A. T. Jotheeswaran, I. Acosta, Z. Liu and M. J. Prince (2018). "The Prevalence and Correlates of Frailty in Urban and Rural Populations in Latin America, China, and India: A 10/66 Population-Based Survey." J Am Med Dir Assoc 19(4): 287-295.e284.
23. McLaughlin, P. M., K. M. Sunderland, D. Beaton, M. A. Binns, D. Kwan, B. Levine, J. B. Orange, A. J. Peltsch, A. C. Roberts, S. C. Strother and A. K. Troyer (2021). "The Quality Assurance and Quality Control Protocol for Neuropsychological Data Collection and Curation in the Ontario Neurodegenerative Disease Research Initiative (ONDRI) Study." Assessment 28(5): 1267-1286.
24. Parsa, N., M. J. Zibaeenezhad, M. Trevisan, A. Karimi Akhormeh and M. Sayadi (2020). "Magnitude of the Quality Assurance, Quality Control, and Testing in the Shiraz Cohort Heart Study." Biomed Res Int 2020: 8179795.
25. Peakman, T. C. and P. Elliott (2008). "The UK Biobank sample handling and storage validation studies." Int J Epidemiol 37 Suppl 1: i2-6.
26. Prince, M., D. Acosta, C. P. Ferri, M. Guerra, Y. Huang, J. J. Llibre Rodriguez, A. Salas, A. L. Sosa, J. D. Williams, M. E. Dewey, I. Acosta, A. T. Jotheeswaran and Z. Liu (2012). "Dementia incidence and mortality in middle-income countries, and associations with indicators of cognitive reserve: a 10/66 Dementia Research Group population-based cohort study." Lancet 380(9836): 50-58.
27. Prince, M. J., D. Acosta, M. Guerra, Y. Huang, K. S. Jacob, I. Z. Jimenez-Velazquez, A. T. Jotheeswaran, J. J. Llibre Rodriguez, A. Salas, A. L. Sosa, I. Acosta, R. Mayston, Z. Liu, J. J. Llibre-Guerra, A. M. Prina and A. Valhuerdi (2021). "Intrinsic capacity and its associations with incident dependence and mortality in 10/66 Dementia Research Group studies in Latin America, India, and China: A population-based cohort study." PLoS Med 18(9): e1003097.
28. Stanfill, B. A., E. S. Nakayasu, L. M. Bramer, A. M. Thompson, C. K. Ansong, T. R. Clauss, M. A. Gritsenko, M. E. Monroe, R. J. Moore, D. J. Orton, P. D. Piehowski, A. A. Schepmoes, R. D. Smith, B. M. Webb-Robertson and T. O. Metz (2018). "Quality Control Analysis in Real-time (QC-ART): A Tool for Real-time Quality Control Assessment of Mass Spectrometry-based Proteomics Data." Mol Cell Proteomics 17(9): 1824-1836.
29. Weimers, P., J. Halfvarson, M. C. Sachs, R. Saunders-Pullman, J. F. Ludvigsson, I. Peter, J. Burisch and O. Olén (2019). "Inflammatory Bowel Disease and Parkinson's Disease: A Nationwide Swedish Cohort Study." Inflamm Bowel Dis 25(1): 111-123.
30. Weiskopf, N. G. and C. Weng (2013). "Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research." J Am Med Inform Assoc 20(1): 144-151.
31. Whitney, C. W., B. K. Lind and P. W. Wahl (1998). "Quality assurance and quality control in longitudinal studies." Epidemiol Rev 20(1): 71-80.
32. 顾大男 (2001). "老年健康数据质量评估问题探讨." 人口与经济(02): 38-43.
33. 郭健, 吕浩涵, 李杰, 于伟超, 刘鹏, 顾卫红, 张抒扬 (2021). "中国国家罕见病注册系统架构和数据质量控制及管理流程." 中国数字医学 16(01): 17-22.
34. 杨景丽, 黄文雅, 黄佩瑶, 柳念, 包凯芳, 丁婕, 陈晓亮, 程宁, 郑山, 白亚娜 (2019). "中国队列研究建立和发展现状." 中国公共卫生 35(10): 1393-1399.