· 调查报告与分析 ·

降低死因数据中垃圾编码的应用效果研究：基于区域卫生信息平台的数据分析

桑灏[1]，宋允雪[1]，陈蕾[2]，沈莉[1]，王婉[1]，虞慧婷[2]，徐征宇[3]，刘星航[2]

1. 上海市普陀区疾病预防控制中心，上海 200333；
2. 上海市疾病预防控制中心，上海 201107；
3. 上海市普陀区卫生健康事务管理中心，上海 200333
通信作者：刘星航，E-mail：liuxinghang@scdc.sh.cn；沈莉，E-mail：jiangjiarui331@163.com

【摘　要】目的 利用区域卫生信息平台诊疗数据，探讨其在降低垃圾编码率上的应用价值，从源头提高死因数据质量。**方法** 本研究基于 2023 年 3 月 1 日—2024 年 2 月 29 日上海市普陀区 8 717 例户籍人口死亡审核数据，采用 GBD-2019 中 ICD-10 垃圾编码表筛选根本死因为垃圾编码的数据，并分析其基线特征。利用区域卫生信息平台对垃圾编码数据进行病案诊疗信息补充，由 2 名死因编码员对根本死因进行重新修正或确认，以评估数据质量优化效果。**结果** 数据分析显示，<60 岁人群、居民死亡推断书及死亡地点为“来院已死”“家中”和“其他”的数据，其根本死因的垃圾编码率较高，分别达 14.14%、12.74%、23.81%、13.43% 和 13.04%。基于区域卫生信息平台，垃圾编码诊疗数据匹配成功率为 55.32%，根本死因修正率为 50.00%。经修正后，总垃圾编码率从 7.01% 显著降低至 5.28%（χ^2=22.683，P<0.001），其中≥60 岁人群、居民死亡医学证明书、居民死亡推断书及死亡地点为“医院病房”、“家中”和“敬老院”的数据垃圾编码率均显著下降（均 P<0.05）。在垃圾编码级别方面，修正前级别Ⅰ（严重影响）~Ⅳ（局限影响）的编码率依次为 1.96%、2.90%、1.28%、0.86%（χ^2=336.015，P<0.001），以级别Ⅱ（重大影响）编码为主，占 41.41%。修正后，级别Ⅰ~级别Ⅲ编码率均显著下降（均 P<0.05），26.48% 的级别Ⅱ编码和 29.46% 的级别Ⅲ编码修正为非垃圾编码。级别Ⅱ的高频编码高血压和高血压脑病、级别Ⅲ的肺的其他疾患和其他特指的呼吸性疾患编码率显著下降（χ^2=15.599，P<0.001；χ^2=4.838，P=0.028）。**结论** 区域卫生信息平台在优化死因数据质量时，能够精准覆盖最主要的死亡报告来源，有效降低垃圾编码率，从而对整体数据质量的提升产生实质性影响。

【关键词】 死因监测；垃圾编码；质量控制；区域卫生信息平台；信息化

Effectiveness of reducing garbage codes in mortality data: a regional health information platform-based analysis

SANG Hao[1], SONG Yunxue[1], CHEN Lei[2], SHEN Li[1], WANG Wan[1], YU Huiting[2], XU Zhengyu[3], LIU Xinghang[2] (*1. Shanghai Putuo District Center for Disease Control and Prevention*, *Shanghai* 200333, *China*; *2. Shanghai Center for Disease Control and Prevention*, *Shanghai* 201107, *China*; *3. Shanghai Putuo District Health Affairs Management Center*, *Shanghai* 200333, *China*)
Corresponding authors: LIU Xinghang, E-mail: liuxinghang@scdc.sh.cn; SHEN Li, E-mail: jiangjiarui331@163.com

【Abstract】 Objective To explore the application value of regional health information platform data in reducing garbage code rates and improving data quality at the source. **Methods** We analyzed 8 717 death records of registered residents in Putuo district, Shanghai from March 1, 2023 to February 29, 2024. Using the GBD-2019 ICD-10, we extracted cases with the underlying cause of death being classified as a garbage code and examined their baseline characteristics. Using the regional health information platform, we supplemented medical records for cases with garbage-coded underlying causes of death. Two trained mortality coders then revised the underlying causes to evaluate data quality improvement. **Results** The garbage code rates were higher for underlying causes of deaths among individuals < 60 years (14.14%), cases documented in resident death inference certificates (12.74%), and deaths occurring at locations categorized as dead on arrival (23.81%), at home (13.43%), and others (13.04%). On the basis of the regional health information platform, the successful matching rate for medical records with garbage-coded underlying causes was 55.32%, and the effective correction rate for underlying causes reached 50.00%. After correction, the overall garbage code rate decreased from 7.01% to 5.28% (χ^2 = 22.683, P < 0.001). Notably, significant reductions were observed in the garbage code rates for individuals ≥ 60 years, cases recorded in both official death certificates and death inference certificates, and deaths occurring in hospital wards, at home, and in nursing homes (all P < 0.05). Regarding garbage code

开放获取：CC BY-NC-ND 4.0 DOI：10.11847/zgggws1147281
基金项目：上海市加强公共卫生体系建设三年行动计划学科人才建设项目（2023—2025）（GWVI-11.1-35、GWVI-11.1-31）；上海市“科技创新行动计划”技术标准项目（22DZ2206000）；上海市加强公共卫生体系建设三年行动计划优秀青年人才项目（2023—2025）（GWVI-11.2-YQ07）；上海市卫生健康委员会卫生行业临床研究专项（20224Y0328）；上海市普陀区卫生健康系统科技创新项目（ptkwws202322）
第一作者：桑灏（1988—），主管医师，硕士，研究方向：生命统计。
收稿日期：2025-03-05　修回日期：2025-05-27　录用日期：2025-08-01
利益冲突：不存在　伦理审查：已获取　出版授权：全体作者已与编辑部签署作者贡献声明及版权转让协议

classification, prior to correction, the coding rates for Level I (severe impact) to Level Ⅳ (limited impact) were 1.96%, 2.90%, 1.28%, and 0.86%, respectively (χ^2 = 336.015, P < 0.001), with Level Ⅱ (major impact) codes predominating at 41.41% of cases. Post-correction, the coding rates for Level I to Level Ⅲ reduced (all P < 0.05), with 26.48% of Level Ⅱ codes and 29.46% of Level Ⅲ codes reclassified as non-garbage codes. High-frequency Level Ⅱ codes such as hypertension and hypertensive encephalopathy, as well as Level Ⅲ codes including other diseases of the lung and other specified respiratory conditions exhibited decreases (χ^2 = 15.599, P < 0.001; χ^2 = 4.838, P = 0.028). **Conclusions** The regional health information platform effectively targets major mortality reporting sources and significantly reduces garbage code rates, thus substantially improving the overall data quality in mortality statistics.

【**Keywords**】 mortality surveillance; garbage codes; quality control; regional health information platform; informatization

死因监测是评估区域人群健康状况和卫生保健水平的重要依据，高质量的死因监测数据对政府制定社会经济发展规划、调整人口发展战略、优化资源配置具有关键指导价值[1–4]。日常监测过程中，尽管全球已有超过 100 个国家（包括中国在内）使用了《疾病和有关健康问题的国际统计分类（第 10 次修订本）》（The International Statistical Classification of Diseases and Related Health Problems 10th Revision, ICD-10）编码体系，但受诊断技术、信息记录完整性以及医务人员认知等因素制约，各国家、地区死因数据质量参差不齐[1, 5–6]。垃圾编码作为死因监测数据中的“噪声”，最早由 Murray C J L 和 Lopez A D[7] 于 1996 年提出，描述在死亡登记中记录的某些根本死因无法为死亡原因分析提供有用的信息，其存在或比例过高会导致死亡谱失真和疾病死亡率趋势误判。世界卫生组织（WHO）数据显示，2012—2016 年全球 20 个国家的垃圾编码率差异较大，从 7%～66%[8]。高社会经济发展指数国家通常具有较低的高影响垃圾编码比例，而社会经济发展指数低的国家则比例较高[8]。中国的垃圾编码率在 10%～19% 之间，地区间差异明显[1]。以往研究虽然提出了一些垃圾编码处理方法如专家咨询法、比例分配法、统计建模等[9–11]，一定程度上解决了死因数据偏倚，但因各国、地区疾病模式差异，难以建立统一标准。信息化发展为死因监测提供了新思路，构建区域卫生信息平台，实现信息共享，为实现更精准的死因判断创造了条件[12–13]。本研究旨在采用区域卫生信息平台诊疗数据，回顾性对死因数据中根本死因为垃圾编码的数据进行再处理，优化根本死因，进而评估其在提升死因监测数据质量方面的应用效果。

1 资料与方法

1.1 资料来源　本研究使用“上海市死亡登记信息系统”中统计日期为 2023 年 3 月 1 日—2024 年 2 月 29 日，户籍地为“普陀区”的死亡报告卡审核数据作为全量数据，共计 8 717 例。本研究经上海市疾病预防控制中心医学科研伦理委员会审查批准（批号：KY-2024-6）。

1.2 上海市死亡登记报告及审核流程　上海市死亡证明分为三类：（1）居民死亡医学证明书，由救治医疗机构或急救中心签发；（2）居民死亡推断书，由社区卫生服务中心根据家属申报签发；（3）居民死亡确认书，由公安司法部门签发。所有死亡信息均需通过“上海市死亡登记信息系统”上报。普陀区负责审核本区户籍人口的死亡报告卡，对死因不明或根本死因存疑的个案，通过调阅病史、社区复核等方式对根本死因进行修正，生成审核数据库。

1.3 垃圾编码数据筛选及分析　基于 2019 全球疾病负担研究（Global Burden of Disease Study 2019, GBD-2019）的 ICD-10 垃圾编码表[14]，筛选全量数据中根本死因为垃圾编码的记录，并标记其等级。分析垃圾编码的基线数据特征。GBD-2019 将垃圾编码按政策影响程度分为 4 级：Ⅰ级（严重影响）：真正未分类的疾病可能属于不止一个的广泛原因组，可能误导流行病学趋势判断（如败血症）；Ⅱ级（重大影响）：真正未分类的疾病可能属于 2 个广泛的组别中的 1 个或 2 个的原因，影响疾病类别内主要原因构成（如原发性高血压）；Ⅲ级（重要影响）：真正未分类的疾病可能在同一 ICD-10 章节内（如未指明部位的癌症）；Ⅳ级（局限影响）：真正未分类的疾病局限于同一种疾病或伤害类别（如未指明类型的中风）。

1.4 区域卫生信息平台处理垃圾编码的应用分析　上海市普陀区区域卫生信息平台涵盖门诊、住院、检验检查等多维度数据。对筛选出的垃圾编码死亡个案通过死者身份证号匹配其死亡前 1 年的诊疗记录。由 2 名生命统计员结合原始填报内容及诊疗数据（死亡诊断、既往史、病程录等）重新审核死因链，修正或确认根本死因。通过计算诊疗数据匹配成功率、根本死因修正率、修正前后垃圾编码率的差异，评估平台应用价值。质控过程：（1）组建专家团队，包括上海市疾病预防控制中心流行病学专家、生命统计学专家及临床医学专家；（2）2 名死因编码员完成背对背独立审核编码；（3）对 2 名死因编码员判定不一致的案例，启动复核流程由专家团队讨论确定最终编码。垃圾编码率（%）=垃圾编码例数/审核数据例数×100%。匹配成功率（%）=区域卫生信息平台匹配成功例数/垃圾编码例数×100%。根本死因修正率（%）=

基于匹配的根本死因修正例数/区域卫生信息平台匹配成功例数×100%。

1.5 统计分析 使用 SPSS 25.0 软件及 Excel 进行数据清理、匹配及分析。率(比例)的组间比较采用 χ^2 检验，多组比较使用 Bonferroni 校正进行 P 值调整。检验水准 α=0.05，所有统计检验均为双侧检验。

2 结 果

2.1 死因监测数据基本情况 本研究使用上海市普陀区户籍人群死亡审核数据，总例数 8 717 例，其中男性占 54.76%(4 773/8 717)。平均年龄(78.92±12.87)岁，最小 6 d，最大 108 岁，≥60 岁人群占总死亡数的 94.48%(8 236/8 717)。以居民医学死亡证明书为主，占 78.54%(6 846/8 717)，其次为居民死亡推断书和居民死亡确认书，分别占 20.35%(1 774/8 717)和 1.11%(97/8 717)。死亡地点以“医院病房”和“家中”为主，占总死亡的 72.96%(6 360/8 717)，其次为“急诊室”[13.49%(1 176/8 717)]、“敬老院”[7.84%(683/8 717)]、“外地”[3.93%(343/8 717)]、“其他”[(1.06%(92/8 717)]和“来院已死”[0.72%(63/8 717)]。报告医疗机构归属地为“普陀区”的数据占 57.27%(4 992/8 717)。

2.2 基于区域卫生信息平台诊疗数据匹配对根本死因中垃圾编码的影响(表 1) 不同特征人群，其垃圾编码率呈现不同特征。<60 岁人群的垃圾编码率 [14.14%(68/481)] 高于≥60 岁人群 [6.59%(543/8 236)]，差异有统计学意义(χ^2=39.683，P<0.001)。居民死亡推断书的垃圾编码率 [12.74%(226/1 774)] 显著高于居民死亡医学证明书 [5.46%(374/6 846)] (χ^2=115.192，P<0.001)。死亡地点为“来院已死”[23.81%(15/63)]、“家中”[13.43%(213/1 586] 和“其他”[13.04%(12/92)] 及报告医疗机构属地为“非普陀区”的垃圾编码率 [8.86%(330/3 725] 相对较高 (χ^2=243.944，P<0.001；χ^2=34.146，P<0.001)。依据 GBD—2019 中的 ICD-10 垃圾编码，普陀区户籍人群死亡审核数据中共筛选出垃圾编码数据 611 条。依据个案身份证号于区域卫生信息平台对垃圾编码数据进行诊疗数据匹配，总匹配成功率为 55.32%，其中≥60 岁人群(58.75%)、居民死亡医学证明书(62.03%)、死亡地点为“医院病房”(69.88%)、“急诊”(58.49%)和“其他”(58.33%)的人群匹配成功率较高，而死亡地点为“外地”(23.68%)、报告医疗机构属地为“非普陀区”(37.58%)的数据匹配成功率较低。基于匹配结果，对根本死因进行重新修正，根本死因修正率为 50.00%。根本死因修正前，普陀区户籍人群死亡审核数据中根本死因的垃圾编码率为 7.01%(611/8 717)。对根本死因进行修正后，原垃圾编码发生等级下降或修正为非垃圾编码，垃圾编码率降至 5.28%(460/8 717)，差异有统计学意义(χ^2=22.683，P<0.001)。其中≥60 岁人群、居民死亡医学证明书、居民死亡推断书及死亡地点为“医院病房”、“家中”和“敬老院”的数据垃圾编码率均显著下降(均 P<0.05)。

表 1 垃圾编码分布特征及区域卫生信息平台数据匹配前、后垃圾编码修正情况

Table 1 Distribution of garbage codes and correction of garbage codes before and after data matching with the regional health information platform

项目	分类	审核数据	根本死因修正前垃圾编码		区域卫生信息平台匹配成功		基于匹配的根本死因修正		根本死因修正后垃圾编码		修正前、修正后垃圾编码率比较	
		例数	例数	率/%	例数	率/%	例数	率/%	例数	率/%	χ^2值	P值
性别	男性	4 773	317	6.64	175	55.21	75	42.86	252	5.28	7.881	0.005
	女性	3 944	294	7.45	163	55.44	94	57.67	208	5.27	15.734	<0.001
年龄/岁	<60	481	68	14.14	19	27.94	10	52.63	62	12.89	0.320	0.571
	≥60	8 236	543	6.59	319	58.75	159	49.84	398	4.83	23.697	<0.001
报卡类型	死亡医学证明书	6 846	374	5.46	232	62.03	117	50.43	270	3.94	17.624	<0.001
	死亡推断书	1 774	226	12.74	102	45.13	50	49.02	180	10.15	5.885	0.015
	死亡确认书	97	11	11.34	4	36.36	2	50.00	10	10.31	0.053	0.817
死亡地点	医院病房	4 774	166	3.48	116	69.88	61	52.59	112	2.35	10.804	0.001
	家中	1 586	213	13.43	105	49.30	44	41.90	170	10.72	5.491	0.019
	急诊室	1 176	106	9.01	62	58.49	28	45.16	82	6.97	3.330	0.068
	敬老院	683	61	8.93	33	54.10	24	72.73	40	5.86	4.715	0.030
	外地	343	38	11.08	9	23.68	2	22.22	36	10.50	0.061	0.806
	其他	92	12	13.04	7	58.33	4	57.14	11	11.96	0.050	0.824
	来院已死	63	15	23.81	6	40.00	6	100.00	9	14.29	1.852	0.173
报告医疗机构属地	普陀区	4 992	281	5.63	214	76.16	105	49.07	226	4.53	6.286	0.012
	非普陀区	3 725	330	8.86	124	37.58	64	51.61	234	6.28	17.679	<0.001
合计		8 717	611	7.01	338	55.32	169	50.00	460	5.28	22.683	<0.001

2.3 基于区域卫生信息平台诊疗数据匹配对垃圾编码等级的影响(表 2)　根本死因修正前，级别Ⅰ～级别Ⅳ编码率分别为 1.96%(171/8 717)、2.90%(253/8 717)、1.28%(112/8 717)和 0.86%(75/8 717)，差异有统计学意义(χ^2=336.015，P<0.001)；以级别Ⅱ编码为主，占总垃圾编码的 41.41%(253/611)。根本死因修正后，各级别垃圾编码发生等级下降或修正为非垃圾编码，其中级别Ⅰ～级别Ⅲ的编码率均下降，差异均有统计学意义(均 P<0.05)；26.48%(67/253)的级别Ⅱ编码和 29.46%(33/112)的级别Ⅲ编码修正后变为非垃圾编码。

表 2 区域卫生信息平台数据匹配前、后各等级垃圾编码修正情况
Table 2 Correction of garbage codes at each level before and after data matching with the regional health information platform

类型	根本死因修正前		根本死因修正后的编码变化										根本死因修正后		修正前、修正后垃圾编码率比较	
	垃圾编码		级别Ⅰ		级别Ⅱ		级别Ⅲ		级别Ⅳ		非垃圾码		垃圾编码			
	例数	率/%	例数	%	例数	%	例数	%	例数	%	例数	率/%	例数	率/%	χ^2值	P值
级别Ⅰ	171	1.96	128	74.85	4	2.34	1	0.58	4	2.34	34	19.89	128	1.47	6.292	0.012
级别Ⅱ	253	2.90	—	—	183	72.33	2	0.79	1	0.40	67	26.48	187	2.15	10.156	0.001
级别Ⅲ	112	1.28	—	—	—	—	79	70.54	0	0.00	33	29.46	82	0.94	4.691	0.030
级别Ⅳ	75	0.86	—	—	—	—	—	—	58	77.33	17	22.67	63	0.72	1.052	0.305

注：“—”表示该项无数据。

2.4 基于区域卫生信息平台诊疗数据匹配对高频垃圾编码的影响(表 3)　依据修正前等级内各编码占比，按降序选取占比累计超过 50% 的前若干位编码，定义为级别内的高频编码。级别Ⅰ的高频编码为原因不明确和未特指原因的死亡(R99、R96.1)及衰老(R54)；级别Ⅱ的高频编码(I10)为高血压和高血压脑病(I67.4)；级别Ⅲ的高频编码分别为肺的其他疾患(J98.4)和其他特指的呼吸性疾患(J98.8)、心脏病的并发症和不明确表述(I51)、肠道部位及消化器官其他和不明确的恶性肿瘤(C26.0、C26.9)、支气管扩张(症)(J47)和部位未特指的恶性肿瘤(C80)；级别Ⅳ的高频编码分别为未特指的支气管肺炎(J18.0)和未特指的肺炎(J18.9)、脑卒中(I64)和脑卒中后遗症(I69.4)。根本死因修正后，除部位未特指的恶性肿瘤(C80)外，各等级的高频编码率均有下降。其中，级别Ⅱ的高血压(I10)及高血压脑病(I67.4)的编码率由修正前的 2.24%(195/8 717)显著降至修正后的 1.43%(125/8 717)(χ^2=15.599，P<0.001)，其占总垃圾编码的比例从 31.91%(195/611)降至 27.17%(125/460)。除外，级别Ⅲ的肺部其他疾患和其他特指的呼吸性疾患(J98.4、J98.8)的编码率经修正后亦显著下降(χ^2=4.838，P=0.028)。

表 3 区域卫生信息平台数据匹配前、后各等级高频垃圾编码修正情况
Table 3 Correction of garbage codes in high-frequency codes before and after data matching with the regional health information platform

类型	顺位	根本死因修正前垃圾编码			根本死因修正后垃圾编码			修正后垃圾编码	χ^2值	P值
		ICD-10编码	例数	率/%	ICD-10编码	例数	率/%	例数下降比例/%		
级别Ⅰ	第1位	R99、R96.1	66	0.76	R99、R96.1	50	0.57	24.24	2.222	0.136
	第2位	R54	46	0.53	R54	30	0.34	34.78	3.383	0.066
级别Ⅱ	第1位	I10、I67.4	195	2.24	I10、I67.4	125	1.43	35.90	15.599	<0.001
级别Ⅲ	第1位	J98.4、J98.8	24	0.28	J98.4、J98.8	11	0.13	54.17	4.838	0.028
	第2位	I51	12	0.14	C80	6	0.07	50.00	2.002	0.157
	第3位	J47	12	0.14	C26.0、C26.9	6	0.07	58.33	2.885	0.089
	第4位	C26.0、C26.9	8	0.09	I51	6	0.07	25.00	0.286	0.593
	第5位	C80	6	0.07	J47	5	0.06	0.00	—	—
级别Ⅳ	第1位	J18.0、J18.9	26	0.30	J18.0、J18.9	15	0.17	42.31	2.958	0.085
	第2位	I64、I69.4	25	0.29	I64、I69.4	20	0.23	20.00	0.557	0.455

注：R99 其他原因不明确和未特指原因的死亡，R96.1 死亡发生于症状起始后24 h以内，R54 衰老，I10特发性(原发性)高血压，I67.4高血压脑病，J98.4 肺的其他疾患，J98.8 其他特指的呼吸性疾患，I51 心脏病的并发症和不明确表述，C26.0 肠道部位未特指的恶性肿瘤；C26.9 消化器官其他和不明确的恶性肿瘤，J47 支气管扩张(症)，C80 部位未特指的恶性肿瘤，J18.0未特指的支气管肺炎，J18.9 未特指的肺炎，I64 脑卒中，I69.4 脑卒中后遗症；“—”表示该项无数据。

3 讨 论

垃圾编码的识别与纠正是提升数据质量、优化公共卫生决策的核心，如何在保证数据不失真的情况下，尽可能地降低垃圾编码率，还原真实死亡谱和研判疾病死亡率趋势是死亡监测工作的意义所在。

普陀区户籍人群死亡数据显示，垃圾编码更常见于<60 岁人群、居民死亡推断书以及死亡地点为“来院已死”“家中”和“其他”的个案。这一差异可能与不同人群的死因特征及医疗记录完整性有关。<60 岁人群可能因意外伤害、急性病症未及时就医，死因推断依赖有限信息，易出现症状性编码（如猝死、不明原因），导致垃圾编码率升高[15]。≥60 岁人群多因慢性病（如心血管疾病、糖尿病、慢性阻塞性肺病）死亡，医疗记录较完整，ICD-10 编码准确性较高。居民死亡推断书主要用于家中或非医疗机构死亡，通常由社区医生依据家属口述或有限信息推断死因，易受信息缺失或记忆偏差影响，易出现垃圾编码。既往研究显示，超过 75% 的心血管疾病死亡发生在家中[16]，患呼吸系统疾病、精神病以及神经系统疾病（如癫痫）的患者，在医疗机构死亡的比例反而较低[17–18]。此外，对于来院已死的个案，其约占急诊死亡总数的 61.90%[19]，急诊死亡多以循环系统疾病、恶性肿瘤、呼吸系统疾病和创伤为主[19–20]。该人群根本死因垃圾编码率较高，将会导致相关类别疾病的关键死因被掩盖，影响疾病负担评估。

本研究结果显示，通过匹配区域卫生平台的诊疗数据，对根本死因修正后，垃圾编码率从 7.01% 降至 5.28%，下降显著。提示该平台可系统性修正不准确的死因记录。尽管匹配成功率为 55.32%，但匹配成功数据中 50.00% 的根本死因得到修正，表明即使在数据覆盖有限的情况下，平台仍能实现较高的修正效率。值得注意的是，该方法的修正效果在特定人群和场景中表现尤为显著，包括≥60 岁人群、居民死亡医学证明书、居民死亡推断书、死亡地点为“医院病房”、“家中”和“敬老院”的数据。进一步分析显示，这些关键群体覆盖了户籍死亡人群的绝大部分；≥60 岁人群占总死亡的 94.48%，居民死亡医学证明书和居民死亡推断书占总死亡的 98.89%；而死亡地点为“医院病房”、“家中”和“敬老院”的个案合计占总死亡的 80.80%。这些分布特征提示，区域卫生平台在优化死因数据质量时，能够精准覆盖最主要的死亡报告来源，从而对整体数据质量的提升产生实质性影响。

本研究结果显示，使用区域卫生平台诊疗数据应对各级别的垃圾编码，能使级别Ⅰ～级别Ⅲ的编码率均有显著下降，特别是级别Ⅱ、级别Ⅲ编码。国际研究通常将败血症、心力衰竭、不明原因死亡、肺栓塞及其他未定义的疾病列为级别Ⅰ高频编码[3, 21–22]，与本研究结果存在差异。普陀区户籍人群死亡审核数据中，级别Ⅰ编码的高频编码为不明原因死亡和衰老等，级别Ⅰ编码率（1.96%）显著低于挪威研究报道的主要垃圾编码率（14.1%）和次要垃圾编码率（15.0%）[3]。主要原因是根据上海市疾病预防控制中心业务规范，已对初始填报为败血症、心力衰竭、肺栓塞的个案进行重新查阅病案或复核调查工作，对根本死因进行了修正。但此过程，会耗费大量人力与时间成本，或因家属对病史掌握不足、急诊病史单一而影响数据的完整性和准确性。尽管普陀区总体垃圾编码率（含级别Ⅱ）处于较低水平，但级别Ⅱ编码占比相对较高，是后续改进的重点。本研究中，高血压作为级别Ⅱ的高频编码，其修正前占总垃圾编码的比例（31.91%）略高于上海既往研究（27.70%），修正后降至 27.17%，与上海市数据持平[23]。编码率由修正前的 2.24% 显著降至修正后的 1.43%，但仍高于挪威报道的 0.50%[3]。高血压患者全死因死亡和心脑血管死亡比例分别为 4.60% 和 1.64%[24]，而缺血性卒中和脑出血更易被归类为垃圾编码（如高血压）[23]，故准确填报高血压合并（并发）症尤为重要。高血压作为高频编码可能源于以下因素：医务人员仅填写高血压而未关联合并症（如心力衰竭、肾衰竭等）；诊断信息不完整、死者突发死亡或来院已死，未明确临床诊断，仅能依据既往史推断；ICD-10 编码局限性，高血压虽然为独立编码，但作为根本死因的适用性较低，常被错误归类（如误判为死因链的终点而非中间环节）。本研究基于区域卫生信息平台诊疗数据的补充，能有效降低级别Ⅱ中高频编码高血压相关编码率，对死因链逻辑错误具有纠正能力（如将“高血压”修正为实际根本死因，如脑卒中、冠心病）。除外，对级别Ⅲ中高频编码肺部其他疾患和其他特指的呼吸性疾患修正效果显著，减少临床诊断模糊性导致的垃圾编码（如补充肺部感染等的具体病因）。

本研究也存在一定局限。数据分析显示，非本区医疗机构报告的死亡数据及死亡地点为“外地”的数据，其数据匹配效果欠佳，仅为 37.58% 和 23.68%，可能造成本研究结果在外推过程中存在一定偏倚。其次，部分特定死因如未明确部位的恶性肿瘤，其编码率未见明显下降，这一现象可能反映了基层医疗机构在诊断规范性、临床编码员编码不准确[25] 和死因填报准确性方面仍需

加强。随着区域卫生信息化建设的持续推进，平台数据覆盖面和应用场景正在不断扩展。特别是通过家庭医生签约项目的实施，各级各类医疗机构的诊疗信息将逐步纳入平台系统，数据匹配成功率将进一步提高。

综上所述，区域卫生信息平台诊疗信息的整合降低了普陀区死因监测数据的垃圾编码率，有效提升了数据质量。值得注意的是，该平台并非专为死因判定设计，在应用场景上需审慎。临床医生的专业判断仍是确保居民死亡证明书数据质量的核心要素，住院死亡病例的医学死亡证明开具应以主治医师的诊断为主，避免因过度校正导致数据偏倚。虽然平台提供的诊疗信息能够补充家属对患者病史描述的完整性，包括住院、门急诊及既往病史等，但仍存在一定局限性，无法完全还原患者临终前的症状体征，对于突发伤害事件、慢病共病急发[15]等情况的判断仍存在挑战。

数据可用性声明　本研究的数据涉及公共卫生敏感信息。根据国家相关法律法规和政策要求，数据必须严格保密，不得擅自公开。在研究过程中，研究团队采取了一系列严格的安全措施来保护数据的安全。因此，无法公开原始数据。

参考文献

[1] 章亚男, 宇传华. 死因统计中的垃圾编码及其处理方法 [J]. 公共卫生与预防医学, 2016, 27(2): 8-11.

[2] de Oliveira C M, Ciríaco D L, da Silva C F, et al. The impact of the investigation on deaths classified as garbage codes on the quality of the cause-of-death information in the Northeast region, Brazil[J]. Revista Brasileira de Epidemiologia, 2019, 22(S3): e19003.

[3] Ellingsen C L, Alfsen G C, Ebbing M, et al. Garbage codes in the Norwegian Cause of Death Registry 1996–2019[J]. BMC Public Health, 2022, 22(1): 1301.

[4] Franca E, Ishitani L H, Teixeira R, et al. Changes in the quality of cause-of-death statistics in Brazil: garbage codes among registered deaths in 1996–2016[J]. Population Health Metrics, 2020, 18(S1): 20.

[5] Hong T T, Hoa N P, Walker S M, et al. Completeness and reliability of mortality data in Viet Nam: implications for the national routine health management information system[J]. PLoS One, 2018, 13(1): e0190755.

[6] 任丽君, 周脉耕, 王黎君. 死因登记报告系统现状与发展 [J]. 疾病监测, 2008, 23(1): 1-3, 10.

[7] Murray C J L, Lopez A D. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020[R]. Cambridge: Harvard University Press, 1996.

[8] Iburg K M, Mikkelsen L, Adair T, et al. Are cause of death data fit for purpose? Evidence from 20 countries at different levels of socio-economic development[J]. PLoS One, 2020, 15(8): e0237539.

[9] 刘利群, 万霞. 死因数据中垃圾编码再分配方法研究进展 [J]. 中华流行病学杂志, 2022, 43(5): 784-788.

[10] Naghavi M, Makela S, Foreman K, et al. Algorithms for enhancing public health utility of national causes-of-death data[J]. Population Health Metrics, 2010, 8: 9.

[11] Foreman K J, Naghavi M, Ezzati M. Improving the usefulness of US mortality data: new methods for reclassification of underlying cause of death[J]. Population Health Metrics, 2016, 14: 14.

[12] 中华人民共和国国家卫生和计划生育委员会. 基于居民健康档案的区域卫生信息平台技术规范: WS/T 448—2014[S]. 北京: 中国标准出版社, 2014.

[13] 徐静, 田琪, 高昭昇, 等. 区域卫生信息平台助力统计工作智能化转变的探索与实践 [J]. 中国卫生信息管理杂志, 2024, 21(6): 819-824.

[14] Institute for Health Metrics and Evaluation. Global Burden of Disease Study 2019 (GBD 2019) cause list mapped to ICD codes[R]. Seattle: Institute for Health Metrics and Evaluation (IHME), 2020.

[15] Johnson S C, Cunningham M, Dippenaar I N, et al. Public health utility of cause of death data: applying empirical algorithms to improve data quality[J]. BMC Medical Informatics and Decision Making, 2021, 21(1): 175.

[16] Wang W, Liu Y N, Ye P P, et al. Trends and associated factors in place of death among individuals with cardiovascular disease in China, 2008–2020: a population-based study[J]. The Lancet Regional Health Western Pacific, 2022, 21: 100383.

[17] 景军, 袁兆宇. 在医院去世与在家中去世——有关中国公民死亡地点的社会学辨析 [J]. 思想战线, 2016, 42(2): 14-18.

[18] 闫范书, 余灵灵, 王黎君, 等. 2021 年中国居民癫痫死亡地点及影响因素分析 [J]. 中国预防医学杂志, 2024, 25(5): 513-518.

[19] 倪伟, 刘亚, 金学敏, 等. 588 例急诊死亡病例的回顾性分析 [J]. 临床急诊杂志, 2020, 21(8): 667-670.

[20] 吕运秀, 孟海英. 2378 例急诊死亡病例患者疾病构成及帕累托图分析 [J]. 中国病案, 2025, 26(1): 65-68.

[21] Naghavi M, Richards N, Chowdhury H, et al. Improving the quality of cause of death data for public health policy: are all 'garbage' codes equally problematic?[J]. BMC Medicine, 2020, 18(1): 55.

[22] 俞浩, 范周全, 罗鹏飞, 等. 江苏省居民归因于代谢异常的心血管病疾病负担分析 [J]. 中华流行病学杂志, 2018, 39(12): 1596-1601.

[23] Chen L, Xia T, Yuan Z A, et al. Are cause of death data for Shanghai fit for purpose? A retrospective study of medical records[J]. BMJ Open, 2022, 12(2): e046185.

[24] 刘仕俊, 张艳, 姜彩霞, 等. 高血压患者可干预危险因素与全死因死亡及心脑血管死亡关系研究 [J]. 中国公共卫生, 2023, 39(9): 1102-1107.

[25] 胡颖辉, 闫学丽, 孔德莹, 等. 原发灶不明恶性肿瘤的 ICD-10 编码质量分析 [J]. 中国病案, 2025, 26(2): 52-55.