Cardiac Analytics and Innovation（CardiacAI）数据存储库：用于转化心血管研究的澳大利亚数据资源

Victoria Blake[1], [2*]、Louisa Jorm[2], Jennifer Yu[1,3], Astin Lee[4,5], Blanca Gallego[2]、Sze- Yuan Ooi[1], [3**]**

[1]东部心脏诊所，威尔士亲王医院，悉尼，新南威尔士州，澳大利亚[2]健康大数据研究中心，新南威尔士大学，悉尼，澳大利亚[3]威尔士亲王临床学院，新南威尔士大学，悉尼，新南威尔士州，澳大利亚[4]医学院，科学、医学与健康学院，伍伦贡大学，新南威尔士州，澳大利亚

[5]The Wollongong Hospital，伍伦贡，新南威尔士州，澳大利亚

*通讯作者：Victoria Blake v.blake@unsw.edu.au

**作者注：Blanca Gallego 和 Sze-Yuan Ooi 是本文的共同高级作者。Blanca Gallego 是高级技术作者，Sze-Yuan Ooi 是高级临床作者。

关键词： 心血管疾病、数据资源、电子病历、转化研究、数据链接、数据共享

背景：

医疗服务提供者每天都会生成海量电子临床数据，涵盖急性医疗、初级医疗和社区医疗环境[1]。此外，个人技术的普及也正在从智能手机健康应用程序以及智能手表等可穿戴设备中产生以往无法获取的健康信息来源[2]。创建能够捕捉完整患者就医历程的综合数据集，是卫生研究人员、政府和临床医生的共同愿景，具有众多潜在应用，可用于改善医疗服务并产生新的洞见[3]。

纵向患者级关联数据对于学习型医疗保健系统的应用至关重要，该系统通过持续评估和改进医疗服务，将研究融入临床实践[3-6]。云计算、机器学习和自然语言处理等关键技术的最新进展，使数据能够实现实时访问、关联和分析。数据科学家、临床医生、政策制定者和患者之间的合作，将确保这一创新与评估循环能够识别尚未满足的需求，提高医疗服务的效率和质量，并迅速评估创新成果，以降低实施过程中的风险和障碍[6]。

慢性疾病（如心血管疾病（CVD））是将研究成果整合到临床实践中的理想重点[3]。CVD 是澳大利亚的首要死因，也是医院支出的主要来源，并对人群造成重大影响，是按伤残调整生命年（DALY）计算的首要疾病负担来源[7-9]。即使只是小幅提升医疗服务效率和患者结局，也能产生重大的社会和经济影响[10]。高效提供心血管护理的一大障碍是医疗卫生系统的碎片化，多个服务提供者分散提供护理，遍及.....

不同学科和应用场景[3,10]。临床医生越来越需要从多个来源收集和解读大量信息，以提供持续且高质量的医疗服务，同时还面临着提高效率的压力[11]。前瞻性地整合健康数据源，将有助于促进多学科协作，并开发能够帮助临床医生更高效地管理心血管疾病（CVD）的分析工具。这反过来将推动高质量、循证且以患者为中心的医疗服务的提供[10,12,13]。

实现这一目标面临着重大挑战。与英国和斯堪的纳维亚等实行全民医疗保健的国家不同，澳大利亚医疗保健数据彼此孤立，鲜少互通。澳大利亚的公立和私立医疗服务提供者构成了复杂的体系，且缺乏统一的电子病历（EMR）系统[10,14]。在医疗服务提供者之间开展此类大规模数据链接时，必须满足诸多监管、数据安全和数据治理要求[3,15]。数据保管人和伦理委员会谨慎审批此类数据链接是可以理解的，因为他们有责任确保患者隐私得到保护、研究和质量改进活动受到适当监督，并确保数据用于造福患者[3,16]。此外，建立安全的数据环境需要投入大量资源和资金，而小型研究团队往往难以获得这些资源和资金。

健康数据关联还面临相当大的技术挑战。电子病历系统往往无法与其他系统连接并交换信息，而且需要使用专有软件才能从中提取信息[17]。一次住院收治的信息可能存储在许多彼此独立的数据库中，而这些系统之间的数据共享有限。此外，电子病历系统并非为研究而设计，因此不同系统中的相似数据类型并未采用标准化格式存储[15]。此外，还存在

实际使用数据时也面临挑战。健康数据包含多种不同类型的数据，例如自由文本、图像和数值测量指标，要有效利用这些数据，需要专业技能、工具和专业知识[18]。即使克服了这些技术和治理方面的障碍，数据及数据提取流程也很少共享，导致研究团队为提取相似信息而重复开展工作，造成效率低下和资源浪费[19]。

CardiacAI 数据存储库

Cardiac Analytics and Innovation（CardiacAI）资料库是一个独立的非营利平台，为澳大利亚研究团队提供安全且受到适当治理的环境，以访问和分析实时心血管电子病历数据。该项目由来自悉尼东南地方卫生区和伊拉瓦拉肖尔黑文地方卫生区（LHD）的资深心脏病专科医生，以及来自新南威尔士大学（UNSW）健康大数据研究中心（CBDRH）的数据科学家共同发起，旨在构建服务于心血管护理的数字医疗与研究生态系统。该项目通过提供所需的研究基础设施，并与数据贡献者建立牢固、互惠互利的合作关系，致力于支持创新的转化研究，从而改善心血管医疗服务提供及其结局。

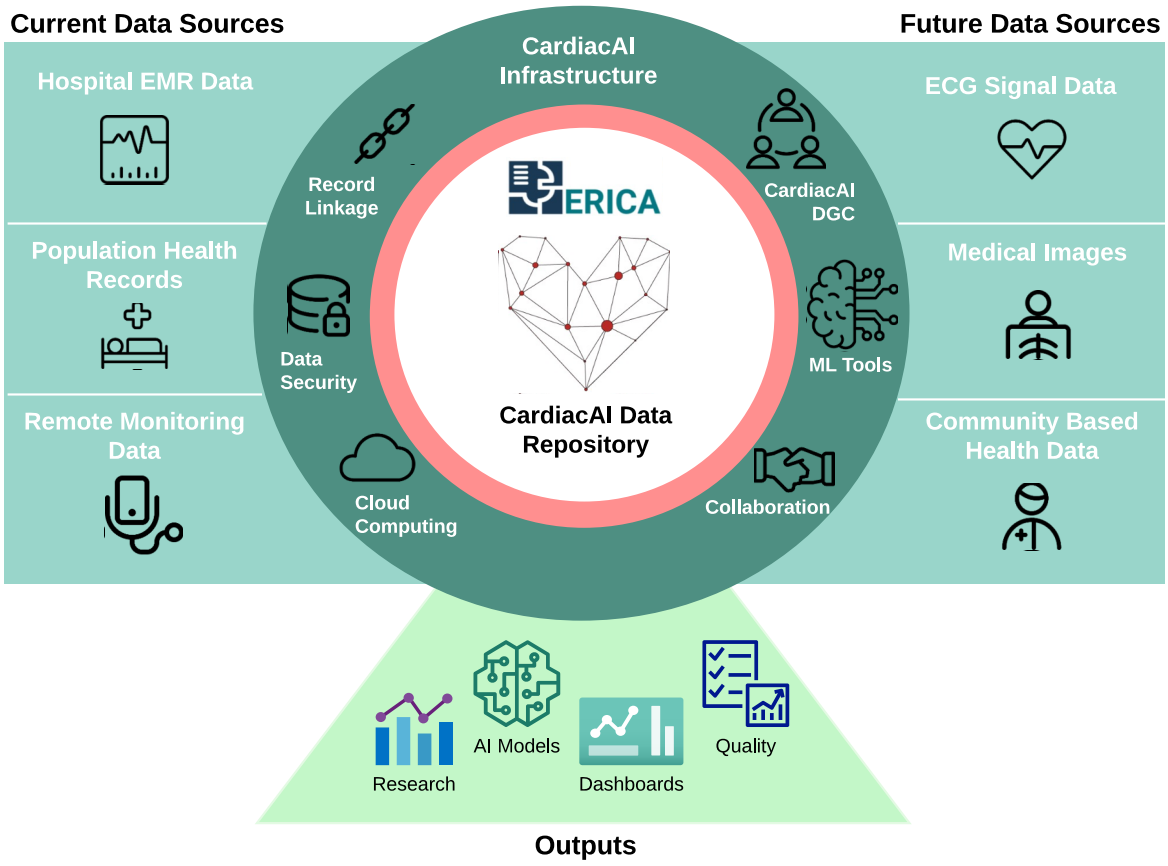


图*1*。*CardiacAI* 项目

CardiacAI 项目前瞻性地关联心血管电子病历数据及其他健康数据源，并将其安全存储于云端托管的安全环境（电子研究机构云架构，ERICA）中；该环境专为存储和使用敏感大型健康数据而设计。该项目目前收集来自悉尼东南地方卫生区（SESLHD）和伊拉瓦拉肖尔黑文地方卫生区（ISLHD）的数据，这两个地方卫生区是澳大利亚新南威尔士州（NSW）15 个地方卫生区中的两个。目前正在扩展数据收集范围，纳入另外两个地方卫生区站点以及一组心血管远程医疗数据集。

为提取医院电子病历数据，团队开发了定制数据库提取脚本，可轻松与采用 Cerner Millennium[20] 电子病历系统的其他澳大利亚医院共享；该系统是澳大利亚公立医院使用的主流系统[21]。

根据需要，将开发适用于其他电子病历和健康数据系统的提取方法，以便逐步将数据源添加到项目中。

该队列包括在四家三级转诊医院接受心血管专科医生收治的所有住院患者，并根据以下标准进行定义：

1) 入院至威尔士亲王医院、St George Public Hospital、The Sutherland Hospital 或 The Wollongong Hospital

2) 在心脏病专科医生、心胸外科医生或血管外科医生管理下入院

3) 入院日期在2017年1月1st 日或之后

符合这些标准的电子病历记录会由提取脚本每晚提取。脚本会获取去标识化电子病历数据，其中包括编码数据字段和数值等结构化信息，以及自由文本临床文档等非结构化数据。

复杂的编码电子病历数据会自动转换为适用于研究的格式，由标准表（例如患者、就诊、诊断和用药管理表）组成，并具有定义明确的主键和外键；必要时，数据元素采用标准本体（例如 SNOMED CT）。

这些电子病历记录采用一种称为概率链接的技术与其他健康数据集进行关联，即根据记录指向同一人的概率进行匹配。NSW Centre for Health Record Linkage (CHeReL) 每年都会与全州人口健康数据集进行关联，包括住院患者数据收集 (APDC)、急诊科数据收集 (EDDC)、NSW出生、死亡和婚姻登记处 (RBDM) 以及死亡原因单条记录文件 (COD-URF)[22]。这些数据集的数据字典链接可在补充材料中找到。

这些行政数据来源可用于确定纳入 CardiacAI 的患者的纵向结局，包括住院、急诊就诊和死亡率。

CardiacAI 伦理与治理框架和数据共享

该项目已获得SESLHD 人体研究伦理委员会（人类研究伦理委员会）（2019/ETH12625）和人口健康服务研究伦理委员会（2020/ETH01614）的伦理批准。该项目采用选择退出式知情同意模式，通过张贴在心血管病房和急诊科的宣传海报向患者介绍项目。资料库中的数据及使用这些数据开展的研究活动均由CardiacAI 数据治理委员会（CardiacAI-DGC）管理。该委员会由健康数据科学家、资深心血管临床医生和数据保管人组成。委员会负责评估数据访问申请，确保数据存储库的完整性与安全性，并确保项目活动遵循经伦理批准的项目方案、数据治理框架以及与数据贡献站点签订的各项协议。该模式使数据贡献者和数据保管人能够继续掌控数据、保护患者和临床医生的隐私，并以有意义且切合实际的方式指导数据使用，从而对心血管护理产生积极影响。

数据访问仅以成本回收方式提供，从而避免重复 CardiacAI 研究团队开展的耗时数据提取和数据关联流程，同时确保资料库的可持续发展与改进。为确保数据访问符合项目的主要目标和伦理标准，拟使用 CardiacAI 数据存储库的研究人员必须提交项目

提案，供 CardiacAI-DGC 审核。提案将根据五安全框架[23] 进行评估，并且必须符合 CardiacAI 项目方案的目标，或适合通过修订方案实现；相关修订将由相应的人类研究伦理委员会评估。

研究人员只能通过 UNSW ERICA 访问数据[24], 。该平台是由新南威尔士大学管理的高度安全云平台，并已通过 eHealth NSW 根据其隐私与安全评估框架（PSAF）进行的认证。ERICA 的构建旨在满足对安全性要求最高的组织的需求，并通过严格的访问控制和监督数据进出平台的过程来保障安全。有关 ERICA 所采用数据安全措施的更多信息，请参阅补充材料。不允许导出原始数据，所有研究成果均由项目高级研究人员审核，以确保符合项目的伦理审批要求并遵循统计披露控制原则[25]。

CardiacAI 数据

CardiacAI 数据存储库中提供了种类广泛的去标识化电子病历和人口健康数据，包括诊断和外科手术、去标识化的自由文本临床文档、病理学和医学影像检查、生理观察数据、药物和血液制品输注，以及入院和病房转移的日期与时间。完整的数据字典可在www.CardiacAI.org/data-dictionary查阅，更多关于数据的说明请参阅补充材料。

临床文本是一项极具价值的信息来源，因为其中包含结构化格式中没有的重要临床细节[26]。鉴于这些信息的敏感性，CardiacAI 项目采取了措施，以确保

患者和临床医生的隐私得到保护。结构化个人身份标识符会在数据提取文件导出至 CardiacAI 前被移除，但患者和临床医生姓名及其他个人身份标识符仍可能嵌入自由文本中。为解决这一问题，CardiacAI 项目建立了一套流程，使用基于澳大利亚临床文本开发的定制去标识化工具自动移除文本中的个人身份标识符[27]；该工具具有很高的准确率（F1分数为 90.6 至 97.17）。此外，去标识化后的临床文本与结构化数据分开存储，并受到严格的访问控制。研究人员必须向 CardiacAI-DGC 申请访问临床文本的许可，并提供充分的理由。这种做法减少了能够访问原始文本的人数，从而降低了重新识别的风险，同时仍可促进研究开展。

CardiacAI队列特征

方法

以下表格旨在概述 CardiacAI 数据存储库中的数据。表格中所显示分类特征的分母为该队列的就诊记录总数。我们已将分析结果与提供各主要疾病组人口统计学、合并症和结局信息的类似研究进行了比较。在选择用于比较的研究时，优先选择澳大利亚研究或多中心研究，而非国际研究或规模较小的单中心研究。

队列

从1st 2017年1月到31st 2020年12月，提取了 CardiacAI 数据存储库中的所有医院电子病历入院记录（就诊记录），作为完整关联人群

这一时期的人口健康数据集可供使用，因此我们能够在分析中纳入结局、诊断和操作代码。我们预计将在2023年收到更新数据后的下一次数据链接结果。通过对入院和出院日期及时间以及机构名称进行分层匹配，将就诊记录与 APDC 数据集中的相应记录关联起来。在50,339条就诊记录中，94条因缺少 APDC 中的对应就诊记录而被剔除，2条因重复的电子病历就诊记录而被剔除，另有2条因就诊类型为“器官获取”而被剔除。图2展示了就诊记录的筛选过程及排除情况。

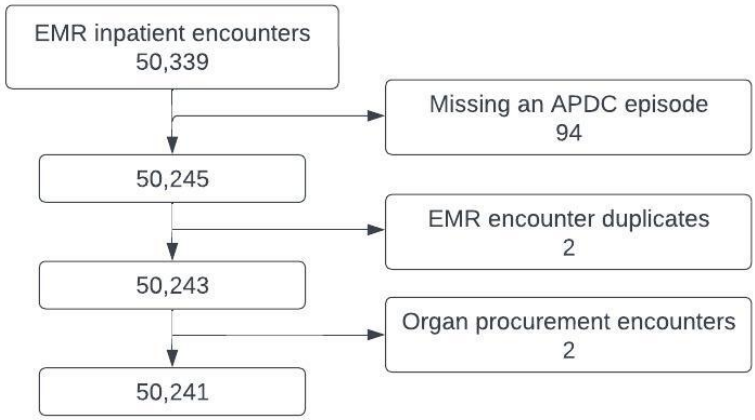


图 *2.* 电子病历就诊记录筛选图。

心血管亚组分析

我们选取了四个心血管亚组，以研究常见心血管疾病患者群体的特征。根据各就诊记录所对应的主要出院诊断代码确定这些群体。各群体及其对应的《国际疾病分类》第十次修订版澳大利亚临床修订版（ICD-10-AM）代码如下：急性冠脉综合征（ACS）I20-25、心力衰竭（HF）I50、心房颤动（AF）I48 和晕厥 R55。亚组按就诊记录层级分类；因此，如果某位患者的就诊记录符合多个亚组的定义，该患者可能会出现在多个组中

。对于不属于这四个组的就诊记录，其主要出院诊断及操作代码见补充材料中的表3。

结局

死亡日期根据 RBDM 和电子病历中的死亡日期，以及 APDC 和 EDDC 中的死亡离院类型确定。RBDM 数据不可用时，则使用电子病历、APDC 和 EDDC 数据。由于死因信息仅提供至 31st 2019年12月，因此我们报告的是全因死亡率，而非心血管死亡。我们预计将在 2023年收到更新的死因数据。

患者因同一事件需要延续护理而在不同医疗机构之间转院的情况很常见，因此，我们通过将 APDC住院事件分组为“就诊事件集”来确定再入院情况。就诊事件集定义为一组连续的就诊记录，并须符合以下条件：1）入院日期为前一次出院日期当天或次日；且 2）前一次就诊记录的离院方式编码为“转院”或“类型变更离院”；或 3）新就诊记录的医疗机构名称与前一次就诊记录中记录的患者转入医疗机构名称一致。

从索引EMR就诊到再入院的天数，是根据索引事件组中最后一次就诊的出院日期，以及下一个相关事件组的首次入院日期计算得出的。就诊事件集中的首次就诊若急诊状态为“急诊”，则判定该再入院是否为非计划再入院。根据ICD-10-AM和澳大利亚

健康干预分类（ACHI）的诊断和操作代码，对主要动脉粥样硬化性、动静脉血栓栓塞性及相关心血管疾病进行分类[28]。使用再入院事件组中首次就诊的诊断和操作代码确定其心血管状态。定义了以下复合结局类别：1）全因非计划再入院或全因死亡（ACread|ACmort）；2）伴有心血管主要或次要诊断或操作，或全因死亡的非计划再入院（CVread|ACmort）；以及3）伴有心血管主要诊断或操作，或全因死亡的非计划再入院（PCVread|ACmort）。所有出院后死亡数据均不包括住院期间死亡。

与既往研究的结果比较

CardiacAI队列特征

CardiacAI 队列包含 2017年1月1st至 2020年12月31st 期间的 50,241 条就诊记录（涉及 36,861 名患者）。表1总结了队列特征。除语言字段仅有三个缺失值外，所分析字段均无缺失值。中位年龄为 71 岁，大多数患者（81.2%）入院时年龄超过 55 岁。男性占 60.5%。1.8%的就诊记录报告了原住民和托雷斯海峡岛民身份，低于全州报告的心血管疾病住院收治比例（2.7%）[29]。在大多数就诊记录中，患者被记录为出生于澳大利亚（59.2%），且主要语言为英语（86%）。四分之三的就诊记录（76.5%）由心脏病专科医生负责。新冠肺炎发生前的三年（2017年至2019年），就诊记录数量保持稳定。然而，2020年的就诊记录数量较新冠肺炎发生前几年的平均就诊记录数量减少了约 700 条（5.5%）。这一模式与此前报告的情况一致，即自新冠肺炎大流行开始以来，心脏就诊数量有所下降，并可能归因于暂停非紧急择期手术以及患者因担心感染新冠肺炎而延迟医院就诊，

此前已有报告显示，自新冠肺炎大流行开始以来，心脏就诊数量有所下降；这可能归因于暂停非紧急择期手术，以及患者因担心感染新冠肺炎-19[30-32]而延迟医院就诊。

大多数（65.1%）就诊记录被归类为急诊（非计划）入院，平均住院时长（LOS）为4.2天（+/- 6.5天）。平均住院时长略低于澳大利亚健康与福利研究所（AIHW）报告的澳大利亚CVD患者平均住院时长（5.9天）[8]。不过，这可能与CardiacAI队列的定义方式有关。CardiacAI纳入了所有由血管外科医生收治的患者，因此该数据集包含非心血管操作，例如静脉曲张手术，其相关住院时间较短。此外，CardiacAI目前不包括通常住院时间较长的卒中患者[8]。最常见的主要诊断代码为I48.9——“未特指的心房颤动和心房扑动”（8.7%），最常见的主要操作代码为38218-00——“冠状动脉造影联合左心导管检查”（10.5%）。

不到十分之一的患者（8.6%）入住重症监护病房（ICU），1.5%的就诊记录发生了院内死亡。出院后六个月内，ACread|ACmort、CVread|ACmort 和 PC Vread|ACmort 的非计划再入院或死亡率分别为31.6%、18.2%和13.8%。大多数非计划再入院（87%）发生在 SESLHD 和 ISLHD 的医院。不过，CardiacAI 仅收录了这些再入院病例中约四分之一（27%）的电子病历数据。这一差异可能是因为患者经其他亚专科诊治后再入院，或再入院至 SESLHD 和 ISLHD 以外的医疗机构。这凸显了关联数据的价值：它能够捕获来自其他专科、其他地方卫生区的医院或私立医疗机构的再入院事件。

表*1 CardiacAI*队列特征

测量指标		结果
患者		36861
就诊记录		50241
专科	心脏病学	38433 (76.5%)
	血管外科	8304 (16.5%)
	心胸外科	3504 (7.0%)
入院 Year	2018	12802（25.5%）
	2019	12795（25.5%）
	2017	12612（25.1%）
	2020	12032（23.9%）
Sex	Male	22298（60.5%）
Age	均值（标准差）	68.5 (15.3)
	中位数 [最小值、最大值]	71.0 [14.0,106.0]
	<45	4097 (8.2%)
	45-54岁	5332 (10.6%)
	55-64岁	9285（18.5%）
	65-74岁	12921（25.7%）
	>=75	18606 (37.0%)
原住民身份	ABTSI	657 (1.8%)
出生国家出生	澳大利亚	21829 (59.2%)
语言	NESB	7020 (14.0%)
长度 Stay	均值（标准差）	4.2（6.5）
	中位数 [最小值, 最大值]	2.1 [0.0,293.5]
	<= 24 小时	13130（26.1%）
	1-3天	17155 (34.1%)
	3-7天	11118 (22.1%)
	7-14天	5961 (11.9%)
	>14 天	2877（5.7%）
入院 Type	急诊	32688（65.1%）
结局 - 住院期间	入住重症监护病房	4299（8.6%）
	全因死亡	740（1.5%）
结局 - 30天	全因死亡	625 (1.3%)
	全因再入院或死亡 (ACread|ACmort)	6506 (13.1%)
	心血管疾病再入院或死亡 (CVread|ACmort)	3663 (7.4%)
	心血管疾病主要诊断再入院或死亡（PCVread|全因死亡）	2655（5.4%）
结局 - 6个月	全因死亡	2195（4.4%）
	全因再入院或死亡 (ACread|ACmort)	15655（31.6%）
	心血管疾病再入院或死亡（心血管疾病再入院|全因死亡）	9004 (18.2%)
	心血管疾病主要诊断再入院或死亡（主要心血管疾病再入院|全因死亡）	6854（13.8%）
主要诊断诊断	I48.9 - 未特指的心房颤动和心房扑动	4365 (8.7%)
	R07.4 - 未特指的胸痛	3786 (7.5%)

	I21.4 - 急性心内膜下心肌梗死	3786（7.5%）
	I25.11 - 自体冠状动脉的动脉粥样硬化性心脏病	3075（6.1%）
	I50.0 - 充血性心力衰竭	3042（6.1%）
主要诊断操作	38218-00 - 冠状动脉造影联合左心导管检查	3858（10.5%）
	38215-00 - 冠状动脉造影	3554（9.7%）
	95550-03 - 辅助医疗干预，物理治疗	2700（7.3%）
	95550-09 - 辅助医疗干预，药学	2167（5.9%）
	38306-00 - 经皮插入 1 个腔内支架至单条冠状动脉	2087（5.7%）

急性冠脉综合征

表2显示了主要诊断为急性冠状动脉综合征（ACS）的就诊记录特征。ACS是CardiacAI中规模最大的心血管亚组，占所有就诊记录的四分之一。男性患者的比例（70.4%）与全州数据[33]一致。不过，CardiacAI中的ACS队列较新南威尔士州（NSW）整体ACS患者年轻：75岁及以上患者占30.1%，而全新南威尔士州为35% [33]。ACS患者入住重症监护病房的比例高于CardiacAI总体人群。入住重症监护病房的数据包括择期和非计划入住，也包括因心脏手术后恢复而入住重症监护病房的患者。出院后的非计划再入院率和死亡率与其他研究[34]报告的水平相似。排名前五的次要出院诊断包括高血压（50.5%）、烟草使用史（27.6%）、当前吸烟（17.0%）和2型糖尿病（15.4%）。

表2 急性冠脉综合征就诊特征

测量指标		**结果**
患者		10818
就诊记录		12518
Sex	Male	8814 (70.4%)
Age	均值（标准差）	68.2（12.4）
	中位数 [最小值, 最大值]	69.0 [19.0、103.0]
	<45	492（3.9%）
	45-54岁	1569（12.5%）

	55-64岁	2993（23.9%）
	65-74岁	3691（29.5%）
	>=75	3773（30.1%）
原住民身份	ABTSI	334（2.7%）
出生国家出生	澳大利亚	7201 (57.5%)
语言	NESB	1598 (12.8%)
长度 Stay	均值（标准差）	4.6 (6.1)
	中位数 [最小值，最大值]	2.8 [0.0,143.6]
	<= 24 小时	2693 (21.5%)
	1-3天	3979 (31.8%)
	3-7天	3408 (27.2%)
	7-14天	1671 (13.3%)
	>14 天	767 (6.1%)
入院 Type	急诊	7811 (62.4%)
结局 - 住院期间	入住重症监护病房	1699（13.6%）
	全因死亡	238（1.9%）
结局 - 30天	全因死亡	105（0.9%）
	全因再入院或死亡（ACread|ACmort）	1508（12.3%）
	心血管疾病再入院或死亡（CVread|ACmort）	882（7.2%）
	心血管疾病主要诊断再入院或死亡（PCVread|ACmort）	651 (5.3%)
结局 - 6个月	全因死亡	351 (2.9%)
	全因再入院或死亡 (ACread|ACmort)	3413（27.8%）
	心血管疾病再入院或死亡（心血管疾病再入院|全因死亡）	1900（15.5%）
	心血管疾病主要诊断再入院或死亡（主要心血管疾病再入院|全因死亡）	1475（12.0%）
前5名次要诊断	I25.11 - 自体冠状动脉动脉粥样硬化性心脏病	6584（54.0%）
	U82.3 - 高血压	6159（50.5%）
	Z86.43 - 烟草使用障碍个人史	3361（27.6%）
	Z72.0 - 当前烟草使用	2078（17.0%）
	E11.9 - 不伴有并发症的2型糖尿病	1879（15.4%）
前5名操作	38215-00 - 冠状动脉造影	2529 (23.4%)
	38218-00 - 冠状动脉造影联合左心导管检查	2526 (23.3%)
	38306-00 - 经皮插入 1 个腔内支架至单条冠状动脉动脉	1981（18.3%）
	38500-00 - 使用 1 个左乳内动脉移植物进行冠状动脉旁路移植术	707（6.5%）
	38306-01 - 经皮插入 2 个或更多腔内支架至单条冠状动脉	573 (5.3%)

心力衰竭

表3显示了主要诊断为心力衰竭的就诊记录特征。心力衰竭就诊记录占CardiacAI就诊记录的比例明显较低（6.9%）。与其他亚组相比，这些患者更可能为男性（57.7%），中位年龄更高（81岁），且出生于澳大利亚以外地区的人员比例更高（53.1%）。极少数就诊记录属于择期就诊，其中92.5%为非计划（急诊）就诊。院内死亡率显著低于其他澳大利亚研究报告的水平（占就诊记录的3.4%）。新南威尔士州和澳大利亚首都领地心力衰竭概况报告称，2013,[35] 院内死亡率为6%；一项使用关联行政数据的研究发现，2010-2015[36]年间院内死亡率从7.9%降至5.1%。这可能表明，越来越多的患者被转至临终关怀机构或回家接受临终照护，而不是在医院死亡。然而，本研究中的30天死亡率为4.3%，与Labrosciano等人[36] 报告的出院后4.1%相近，这可能表明对急性心力衰竭发作患者的识别和治疗已有所改善。30天ACread|ACmort率为24.5%，仅计算全因再入院时为20.2%，与其他澳大利亚研究[34]报告的17%至27%相近。

表3 心力衰竭就诊特征

测量指标		**结果**
患者		2689
就诊记录		3484
Sex	Male	2010（57.7%）
Age	均值（标准差）	78.8（11.3）
	中位数 [最小值、最大值]	81.0 [21.0、103.0]
	<45	53（1.5%）
	45-54岁	101（2.9%）
	55-64岁	272（7.8%）

	65-74岁	635（18.2%）
	>=75	2423（69.5%）
原住民身份	ABTSI	38（1.1%）
出生国家	澳大利亚	1635（46.9%）
语言	NESB	885（25.4%）
时长 Stay	均值（标准差）	6.3（7.6）
	中位数 [最小值，最大值]	4.4 [0.0,293.5]
	<= 24 小时	204（5.9%）
	1-3天	854（24.5%）
	3-7天	1386（39.8%）
	7-14天	741 (21.3%)
	>14 天	299 (8.6%)
入院 Type	急诊	3224 (92.5%)
	入住重症监护病房	125 (3.6%)
结局 - 住院期间	全因死亡	120 (3.4%)
结局 - 30天	全因死亡	143 (4.3%)
	全因再入院或死亡 (全因再入院|全因死亡)	824 (24.5%)
	心血管疾病再入院或死亡（心血管疾病再入院|全因死亡）	634 (18.8%)
	心血管疾病主要诊断再入院或死亡（主要心血管疾病再入院|全因死亡）	485 (14.4%)
结局 - 6个月	全因死亡	517（15.4%）
	全因再入院或死亡 (ACread|ACmort)	1855（55.1%）
	心血管疾病再入院或死亡（心血管疾病再入院|全因死亡）	1489（44.3%）
	心血管疾病主要诊断再入院或死亡（主要心血管疾病再入院|全因死亡）	1221（36.3%）
前5名次要诊断	U82.3 - 高血压	1800 (52.2%)
	U82.1 - 缺血性心脏病	1228（35.6%）
	Z86.43 - 烟草使用障碍个人史	866（25.1%）
	N17.9 - 急性肾衰竭	843 (24.4%)
	I48.9 - 未特指的心房颤动和心房扑动	792 (23.0%)
前5名操作	95550-03 - 辅助医疗干预，物理治疗	791（29.7%）
	95550-09 - 辅助医疗干预，药学	338 (12.7%)
	92209-00 - 无创通气支持管理，24 小时或 less	178（6.7%）
	95550-01 - 辅助医疗干预，社会工作	157 (5.9%)
	38218-00 - 冠状动脉造影联合左心导管检查	148 (5.6%)

心房颤动

表4显示了主要诊断为心房颤动的就诊记录特征。心房颤动就诊记录占所有就诊记录的10.8%，男性比例（55.7%）、中位年龄（72岁）和中位住院时间（1.4天）与其他心房颤动住院治疗研究[37,38]相近。心房颤动队列在所有随访时间段的死亡率均较低（住院期间为0.2%，30天时为0.7%，180天时为2.3%）。非计划再入院或死亡结局率为

PCVread|ACmort在30天和180天时分别为7.9%和18.8%。ACread|ACmort在30天时为13%，在6个月时为31.5%；若仅计算全因再入院，则分别为12.1%和29%。心脏复律是最常见的主要操作，发生于35.7%的所有心房颤动就诊记录中；7.2%的就诊记录中植入了心脏起搏器。全因30天再入院率高于Woods等人报告的结果（9.9%）[38]。心脏复律率也高于Weber等人报告的比率（29.5%）[39]。这些差异可能是由于两项研究采用的队列选择不同所致；两项研究均仅纳入首次心房颤动住院，这可能与较低的再入院率和心脏复律率有关[38,39]。排名靠前的次要诊断包括常见的心血管风险因素和疾病。有趣的是，骨关节炎也位列前五位次要诊断之中，在12.1%的心房颤动就诊记录中有记录。

表 4 心房颤动就诊特征

测量指标		**结果**
患者		4297
就诊记录		5422
Sex	Male	3021 (55.7%)
Age	均值（标准差）	70.0 (13.9)
	中位数 [最小值, 最大值]	72.0 [16.0,103.0]
	<45	302 (5.6%)
	45-54岁	505（9.3%）
	55-64岁	973（17.9%）
	65-74岁	1505（27.8%）
	>=75	2137（39.4%）
原住民身份	ABTSI	55（1.0%）
出生国家出生	澳大利亚	3188 (58.8%)
语言	NESB	756 (13.9%)
长度 Stay	均值（标准差）	2.4（2.9）
	中位数 [最小值，最大值]	1.4 [0.0,50.2]
	<= 24 小时	1581（29.2%）
	1-3天	2541（46.9%）
	3-7天	951（17.5%）
	7-14天	302（5.6%）
	>14 天	47 (0.9%)
入院 Type	急诊	4241 (78.2%)
	入住重症监护病房	27 (0.5%)

结局 - 住院期间	全因死亡	12 (0.2%)
结局 - 30天	全因死亡	36（0.7%）
	全因再入院或死亡（ACread|ACmort）	701（13.0%）
	心血管疾病再入院或死亡（CVread|ACmort）	531（9.8%）
	心血管疾病主要诊断再入院或死亡（主要心血管疾病再入院|全因死亡）	427（7.9%）
结局 - 6个月	全因死亡	127（2.3%）
	全因再入院或死亡（ACread|ACmort）	1704（31.5%）
	心血管疾病再入院或死亡（CVread|ACmort）	1247（23.0%）
	心血管疾病主要诊断再入院或死亡（PCVread|ACmort）	1018（18.8%）
前5名次要诊断	U82.3 - 高血压	2550（52.3%）
	Z86.43 - 烟草使用障碍个人史	1012（20.8%）
	U82.1 - 缺血性心脏病	921（18.9%）
	U86.2 - 关节炎和骨关节炎 [主要]	588（12.1%）
	E11.9 - 无并发症的2型糖尿病	526 (10.8%)
前5名操作	13400-00 - 心脏复律	1276 (35.7%)
	55118-00 - 经食管二维实时超声检查《心脏》	573 (16.0%)
	95550-09 - 辅助医疗干预，药学	382 (10.7%)
	38353-00 - 皮下心脏起搏器发生器植入	257 (7.2%)
	95550-03 - 辅助医疗干预，物理治疗	206 (5.8%)

晕厥

表5展示了主要诊断为晕厥的就诊记录特征。主要诊断为晕厥的就诊记录占CardiacAI就诊记录的3.4%。男性比例（57.7%）和中位年龄（72岁）与其他研究相似[40,41]。大多数就诊记录被归类为急诊（87.5%），且住院时长少于3天（80.4%）。尽管因晕厥住院的患者发生心血管事件和死亡的风险增加，但关于晕厥住院治疗结局的澳大利亚研究仍然有限[42]。近期一项瑞典研究发现，从首次晕厥就诊记录到首次心血管事件的平均时间为3.5年[42]。然而，其他随访时间较短（30天）的国际研究结果与我们的再入院率和死亡率结果一致[41,43]。CVread|ACmort和PCVread|ACmort在30天时分别为2.9%和2.2%，在180天时分别为9.6%和7.1%。在30天和180天时，ACread|ACmort率均为PCVread|ACmort率的2.5倍。一台心脏事件记录器被

在15.3%的晕厥就诊中植入了装置，在9.8%的晕厥就诊中植入了起搏器。

表*5* 晕厥就诊特征

测量指标		结果
患者		1632
就诊记录		1696
Sex	Male	979（57.7%）
Age	均值（标准差）	67.5（19.1）
	中位数 [最小值, 最大值]	72.0 [15.0，106.0]
	<45	239（14.1%）
	45-54岁	148（8.7%）
	55-64岁	207（12.2%）
	65-74岁	385（22.7%）
	>=75	717（42.3%）
原住民身份	ABTSI	10 (0.6%)
出生国家的出生	澳大利亚	1085 (64.0%)
语言	NESB	178 (10.5%)
长度 Stay	均值（标准差）	2.2 (2.3)
	中位数 [最小值，最大值]	1.5 [0.0,23.7]
	<= 24 小时	488 (28.8%)
	1-3天	875（51.6%）
	3-7天	252（14.9%）
	7-14天	69（4.1%）
	>14 天	12 (0.7%)
入院 Type	急诊	1484 (87.5%)
结局 - 住院期间	入住重症监护病房	4（0.2%）
	全因死亡	2（0.1%）
结局 - 30天	全因死亡	10 (0.6%)
	全因再入院或死亡 (ACread|ACmort)	133 (7.8%)
	心血管疾病再入院或死亡 (CVread|ACmort)	49 (2.9%)
	心血管疾病主要诊断再入院或死亡（主要心血管疾病再入院|全因死亡）	37 (2.2%)
结局 - 6个月	全因死亡	40 (2.4%)
	全因再入院或死亡（ACread|ACmort）	408 (24.1%)
	心血管疾病再入院或死亡 (心血管疾病再入院|全因死亡)	163（9.6%）
	心血管疾病主要诊断再入院或死亡（PCVread|全因死亡）	120（7.1%）
前5名次要诊断	U82.3 - 高血压	692 (44.7%)
	U82.1 - 缺血性心脏病	275（17.8%）
	Z86.43 - 烟草使用障碍个人史	247（16.0%）
	U73.9 - 未特指的活动	230（14.9%）
	U86.2 - 关节炎和骨关节炎 [主要]	180（11.6%）
前5名操作	95550-03 - 辅助医疗干预，物理治疗	141（17.8%）
	95550-09 - 辅助医疗干预，药学	122（15.4%）
	38285-00 - 皮下植入监测装置置入装置	121（15.3%）
	38353-00 - 皮下心脏起搏器发生器植入发生器	78 (9.8%)
	95550-01 - 辅助医疗干预，社会工作	34 (4.3%)

讨论

在本研究中，我们介绍了 CardiacAI 数据存储库，这是一个服务于转化心血管研究与创新的宝贵数据资源。通过创建这一可用于研究的前瞻性数据资源库，CardiacAI 项目有望促进协作项目，考察当代与健康相关的活动，描述心血管疾病队列，以务实的方式监测治疗干预和患者结局，发现证据与实践差距，并开发能够满足未满足临床需求的工具。

CardiacAI 项目正在支持多个研究项目，这些项目已累计获得超过 300 万美元的资金支持，包括：1）开发用于心血管患者的医院出院风险分层工具原型及临床仪表板；2）扩展现有的 CardiacAI数据集，纳入卒中患者、远程监测数据以及另外两个地方卫生区的数据；3）开发下一代临床质量登记系统所需的基础设施。

稳健的 CardiacAI基础设施和治理框架是关键的推动因素。CardiacAI-DGC 为数据贡献者和项目团队提供了一个共同管理数据使用的平台。这也营造了协作环境，鼓励本地临床医生和研究人员参与其中，以解答本地临床问题并创建满足本地需求的工具。该项目还通过确保数据由独立方保管，并听取所有相关方的意见，促进公共和私营部门的所有医疗服务提供者参与。此外，ERICA平台严格的数据安全控制措施让数据保管人和患者确信，数据机密性得到了维护。

CardiacAI 项目还通过 ERICA平台促进了人们更便捷地访问详细的电子病历数据。该项目的基础设施和数据存储库不仅为心血管研究提供了重要资源，也为更广泛的澳大利亚健康数据科学提供了重要资源。在开发算法和数据驱动工具时使用本地数据，有望减轻仅基于国际来源数据集开发工具可能产生的算法偏差[44]。这也确保最终的算法能够在澳大利亚患者群体（包括原住民和文化和语言多元群体（CALD）社区）中达到预期表现。

最后，CardiacAI 提取中纳入临床文本，提供了一个超越传统医院行政数据集、内容独特且更加丰富的心血管疾病数据集。CardiacAI项目团队目前正致力于开发临床文本挖掘方法，从非结构化自由文本中提取重要的心血管信息。这项工作是 CardiacAI 项目的重要基石，将为所有项目提供关键数据点，并促进构建稳健的心血管风险评分、表型算法、临床仪表板及其他数据工具[26]。

CardiacAI 的当前局限性与未来发展方向

新的数据源

CardiacAI 数据存储库涵盖为各类多元人群提供服务的地方卫生区的数据，包括来自CALD社区的群体、原住民以及来自不同社会经济背景的个体[45,46]，为新南威尔士州心血管疾病患者提供了丰富的数据。尽管如此，该资料库尚未纳入卒中住院数据（占澳大利亚所有心血管疾病住院治疗的11%）[8],，也未纳入私立医院的数据，而私立医院的心血管疾病住院治疗量接近总量的一半[47]。

此外，CardiacAI 目前缺少心脏导管手术和诊断影像数据，以及初级和社区护理数据集（如全科医疗和心脏康复）——这些数据将为出院期间及出院后的心血管疾病管理提供有价值的见解。

CardiacAI项目团队正在大力推进扩展资料库，使其更具心血管疾病患者群体的代表性。目前，现有队列正在更新，以纳入卒中住院和急诊科死亡。此外，项目正在加入一项基于心血管疾病诊断和手术编码的回顾性电子病历提取数据，其中涵盖患者发生心血管事件、但由非心血管专科医生收治的住院病例。这些更新将确保 CardiacAI 数据涵盖所有发生心脏事件的住院患者。

该项目还将扩展至新南威尔士州的另外两个地方卫生区，其中一个位于新南威尔士州地区，另一个位于悉尼大都市区，为文化和语言多元群体、原住民群体以及不同社会经济群体提供服务。此次扩展将使 CardiacAI 资料库的规模扩大一倍，并解决多样性问题。预计 2023年将获得这些新增入院类型和地点的电子病历数据。此外，其他几个地方卫生区和医疗服务提供者也已表示对该项目感兴趣，项目团队正寻求逐步纳入新的数据源，具体方式包括共享现有的电子病历提取脚本，或调整提取脚本以适配新的医院电子病历系统。

团队还在推进从住院数据以外的其他来源提取数据。面向曾发生急性冠状动脉综合征、心力衰竭或卒中的患者而设计的 TeleClinical Care (TCC)[48] 远程监测应用程序所收集的数据，将于2023年加入资料库。这款创新应用程序正在新南威尔士州的多家医院推广，为社区中的患者提供此前无法获取的定期生物特征观测、用药计划和患者报告指标数据。

此外，团队正在探索纳入心脏导管手术、心电图和心脏成像数据，以及社区数据源，例如病理学、放射学、心脏康复服务、心力衰竭外展服务和全科医疗（GP）数据；这些数据可直接来自全科医生电子病历系统，也可通过与医疗保险福利计划（MBS）和药品福利计划（PBS）等行政数据进行关联获取。

数据质量与标准化

毫无疑问，电子病历数据是快速生成真实世界证据以及创建支持临床决策工具的宝贵资源。然而，众所周知，电子病历数据存在数据质量问题[49,50]。尽管 CardiacAI数据集 已经过广泛测试，以确保其包含的信息与医疗记录中的记录一致，但目前正在对其正确性、完整性、一致性和合理性进行详细且系统的质量评估，相关结果将于近期公布。为促进 CardiacAI 数据用户参与国际研究，我们还计划在该领域既有工作的基础上，将现有的基于澳大利亚的数据标准化流程扩展至观察性医疗结果合作伙伴通用数据模型（OMOP通用数据模型）[51]，，延续该领域此前的工作[52,53]。OMOP通用数据模型是国际上用于支持真实世界证据的主导数据模型，由观察性健康数据科学与信息学（OHDSI）社区维护和使用[54]。

此外，团队还将探索根据从临床文本中提取的概念验证结构化信息，并应用其他成熟工具，以提供更详细的电子病历数据质量评估功能[53,55]。通过采取这些措施，我们旨在

最大限度地发挥CardiacAI数据集的价值，确保其临床信息的可靠性和准确性。

结论

CardiacAI 项目建立了一个独特的心血管数据集，具有巨大潜力，可为预防继发性心血管事件以及改善心血管疾病患者的护理和生活质量带来新的见解与创新。项目团队将依托现有完善的伦理、治理和数据安全基础设施，致力于进一步扩大项目成果，最终目标是将其打造为全面的州级或国家级心血管数据资产。

致谢

我们衷心感谢 SESLHS 和 ISLHD 持续参与 CardiacAI 项目。我们也感谢健康记录关联中心为关联电子病历数据与人口健康数据集所作的贡献。

伦理审批

本项目已获得悉尼东南地方卫生区人类研究伦理委员会（地方参考编号：2019/ETH12625）和新南威尔士州人口健康研究伦理委员会（地方参考编号：2020/ETH01614）的批准。

数据可用性声明

由于澳大利亚隐私立法，无法公开共享本文所依据的数据。研究人员可向 support@cardiacai.org 提交意向书，申请访问 CardiacAI 平台中的数据。有关 CardiacAI 项目数据共享的更多信息，请访问 https://www.cardiacai.org/data。研究代码可在 https://github.com/CardiacAI-AU 获取。

利益冲突

作者声明不存在利益冲突。

资金支持

Cardiac Analytics and Innovation项目是由悉尼东南地方卫生区（SESLHD）、伊拉瓦拉肖尔黑文地方卫生区（ISLHD）与新南威尔士大学（UNSW）健康大数据研究中心（CBDRH）共同开展的协作项目；后者为该项目提供了实物资助。其他资金支持来自2020年医学研究未来基金（MRFF）心血管健康使命资助计划，以及2021年新南威尔士大学医学院心脏、血管与代谢医学创意种子基金计划。本文所表达的观点仅代表作者本人，不代表资助机构。

参考文献

1. Shortreed SM、Cook AJ、Coley RY、Bobb JF、Nelson JC。利用大型医疗保健数据推动医学科学与公共卫生发展的挑战与机遇。《美国流行病学杂志》2019年292 2 188 851 861 101093；：-。doi：。/aje/kwy。Su J、Zhang Y、Ke Q-Q、Su J-K、Yang Q-H。调动人工智能促进心脏远程康复。《心血管医学评论》2022年；23:045。doi：10.31083/j.rcm2302045 3. Maddox TM、Albert NM、Borden WB等。学习型医疗保健系统与心血管护理：美国心脏协会科学声明。《循环》2017年；135:e826-e857。doi：10.1161/CIR.0000000000000480 4. Olsen L、Aisner D、McGinnis JM。学习型卫生系统：研讨会总结。2007年。doi：10.17226/11903 5. Zurynski Y、Smith CL、Vedovi A等。绘制学习型卫生系统：当前证据的范围综述。载于：澳大利亚健康创新研究所与NHMRC卫生系统可持续性合作中心：澳大利亚健康创新研究所、麦考瑞大学；2020年。6. Bhavnani Sanjeev P、Parakh K、Atreja A等。2017年创新路线图——数字健康、大数据与精准健康时代的ACC卫生政策声明：医疗保健转型。《美国心脏病学会杂志》2017年；70:2696-2718。doi：10.1016/j.jacc.2017.10.018 7. 澳大利亚健康与福利研究所。澳大利亚卫生支出2016–17年。
h
t
t
ps://
www.aihw.gov.au/getmedia/e8d37b7d-2b52-4662-a85f-01eb176f6844/aihw-hwe-74.pdf.aspx?inline=true 8. 澳大利亚健康与福利研究所。心脏、卒中和血管疾病：澳大利亚事实。
https://www.aihw.gov.au/reports/heart-stroke-vascular-diseases/cardiovascular-health-compendium 9. 澳大利亚健康与福利研究所。2022年澳大利亚疾病负担研究。
https://www.aihw.gov.au/reports/burden-of-disease/australian-burden-of-disease-study-2022 10. 澳大利亚卫生部长咨询委员会。慢性病国家战略框架：所有澳大利亚人都通过有效预防和管理慢性病过上更健康的生活。
htt
ps://www.health.gov.au/resources/publications/national-strategic-framework-for-chronic-conditions 11. Johnson KW、Torres Soto J、Glicksberg Benjamin S等。心脏病学中的人工智能。《美国心脏病学会杂志》2018年；71:2668-2679。doi：10.1016/j.jacc.2018.03.521 12. Kao DP、Trimkley KE、Lin C-T。电子健康记录赋能的心力衰竭管理创新。《JACC：心力衰竭》2020年；8:223-233。doi：https://doi.org/10.1016/j.jchf.2019.09.008 13. Roca J、Tenyi A、Cano I。诊断范式变化：利用大数据进行预测。《临床化学与实验室医学》2019年；57:317-327。doi：doi:10.1515/cclm-2018-0971

14. Srinivasan U、Rao S、Ramachandran D、Jonas D。《盲目飞行：澳大利亚消费者与数字健康》。https://flyingblind.cmcrc.com/files/files/Flying-Blind--Australian-Consumers-and-Digital-Health.pdf（1）15. Ehrenstein V、Kharrazi H、Lehmann H、Taylor CO。获取电子健康记录中的数据。载于：《注册系统互操作性工具与技术，评估患者结局的注册系统：用户指南》第3版增补版 *2 [*网络*]*：医疗保健研究与质量局（美国）；2019年。16. Allen J、Adams C、Flack F。数据保管人在建立和维护健康研究社会许可方面的作用。《生物伦理学》2019年；33：502-510。doi：10.1111/bioe.12549 17. Denton N、Molloy M、Charleston S等。数据孤岛正在破坏药物开发，并使罕见病患者无法受益。《孤儿药与罕见病杂志》2021年；16。doi：10.1186/s13023-021-01806-4 18. Garmire LX、Gliske S、Nguyen QC等。生物医学领域下一代数据科学家的培养。载于：《生物计算 2017》。2017年，第640-645页。世界科学出版社。19. Chawinga WD、Zinn S。研究数据共享的全球视角：系统性文献综述。《图书馆与信息科学研究》2019年；41：109-122。doi：https://doi.org/10.1016/j.lisr.2019.04.004 20. Cerner。Cerner Millennium。https://www.cerner.com/se/en/solutions/millennium（2021年5月12日）21. Penderson B。澳大利亚的电子病历：国家现状。https://www.hisa.org.au/slides/hic18/tue/BrucePedersen.pdf 22. 新南威尔士州卫生部。健康记录关联中心（CHeReL）。https://www.cherel.org.au/（2021年5月4日）23. Desai T、Ritchie F、Welpton R。五项安全原则：为研究设计数据访问。https://core.ac.uk/download/pdf/323894811.pdf 24. 新南威尔士大学。ERICA——电子研究机构云架构。https://research.unsw.edu.au/erica（2021年5月4日）25. 澳大利亚统计局。《数据保密指南》。https://www.abs.gov.au/about/data-services/data-confidentiality-guide（2022年）26. Percha B。《现代临床文本挖掘：指南与综述》。《生物医学数据科学年度评论》2021年；4：165-187。doi：10.1146/annurev-biodatasci-030421-030931 27. Liu L、Perez-Concha O、Nguyen A、Bennett V、Jorm L。澳大利亚医院出院小结去标识化：使用深度学习模型集成的端到端框架。《生物医学信息学杂志》2022年；135：104215。doi：https://doi.org/10.1016/j.jbi.2022.104215 28. Joshy G、Korda R、Abhayaratna W、Soga K、Banks E。利用常规收集的数据对主要心血管疾病住院治疗进行分类。《公共卫生研究与实践》2015年；25。doi：10.17061/phrp2531532 29. 流行病学与证据中心HN。新南威尔士州心血管疾病住院情况。https://www.healthstats.nsw.gov.au/#/r/106749（2023年2月15日）30. Toner L、Koshy AN、Ko J、Driscoll A、Farouque O。心力衰竭住院治疗的临床特征与趋势：COVID-19 封锁期间的澳大利亚经验。《心力衰竭》2020年；8：872-875。doi：doi：10.1016/j.jchf.2020.05.014 31. Sutherland K、Chessman J、Zhao J等。COVID-19 对澳大利亚新南威尔士州医疗服务活动的影响。《公共卫生研究与实践》2020年；30。doi：10.17061/phrp3042030

32. Cho K、Femia G、Lee R等。COVID-19大流行期间心脏病学患者医院就诊、医疗服务利用及心血管风险因素的探讨。《心脏、肺与循环》2023年。doi：https://doi.org/10.1016/j.hlc.2022.11.013 33. 新南威尔士州流行病学与证据中心HealthStats。冠心病住院治疗。https://www.healthstats.nsw.gov.au/#/r/106753（2023年2月15日）34. Labrosciano C、Air T、Tavella R、Beltrame JF、Ranasinghe I。心血管疾病住院治疗后的再入院：澳大利亚文献的范围综述。《澳大利亚健康评论》2020年；44：93。doi：10.1071/ah18028 35. Newton PJ、Davidson PM、Reid CM等。新南威尔士州和澳大利亚首都领地的急性心力衰竭住院病例：新南威尔士州心力衰竭快照研究。《澳大利亚医学杂志》2016年；204：113-113。doi：10.5694/mja15.00801 36. Labrosciano C、Horton D、Air T等。澳大利亚和新西兰心力衰竭住院治疗后30天全因死亡率和非计划再入院的发生频率、趋势及机构间差异。《欧洲心力衰竭杂志》2021年；23：31-40。doi：10.1002/ejhf.2030 37. Weber C、Hung J、Hickling S等。因心房颤动住院患者中新发心力衰竭的发生率、预测因素及死亡风险。《心脏》2021年；107：1320。doi：10.1136/heartjnl-2020-318648 38. Woods T-J、Ngo L、Speck P、Kaambwa B、Ranasinghe I。澳大利亚和新西兰心房颤动住院治疗后的30天非计划再入院。《心脏、肺与循环》2022年；31：944-953。doi：https://doi.org/10.1016/j.hlc.2022.02.006 39. Weber C、Hung J、Hickling S等。首次心房颤动住院治疗后的紧急再入院和长期死亡风险。《心脏》2023年；109：380。doi：10.1136/heartjnl- 2022-321560 40. Simos P、Scott I。对因晕厥或近晕厥就诊的普通内科患者进行检查时经胸超声心动图的合理使用。《研究生医学杂志》2022年：postgradmedj-2021-141416。doi：10.1136/postgradmedj-2021-141416 41. Sandhu RK、Sheldon RS、Savu A、Kaul P。2004年至2014年晕厥住院治疗及结局的全国趋势。《加拿大心脏病学杂志》2017年；33：456-462。doi：https://doi.org/10.1016/j.cjca.2016.11.005 42. Yasa E、Ricci F、Magnusson M等。原因不明的晕厥和直立性低血压住院治疗后的心血管风险。《心脏》2018年；104：487-493。doi：10.1136/heartjnl-2017-311857 43. Abulhamayel A、Savu A、Sheldon RS、Kaul P、Sandhu RK。加拿大成年人晕厥住院治疗中合并症负担和结局的地理差异。《加拿大心脏病学杂志》2018年；34：937-940。doi：https://doi.org/10.1016/j.cjca.2018.04.011 44. Gianfrancesco MA、Tamang S、Yazdany J、Schmajuk G。使用电子健康记录数据的机器学习算法中的潜在偏倚。《美国医学会杂志·内科学》2018年；178：1544。doi：10.1001/jamainternmed.2018.3763 45. 南东部悉尼地方卫生区。南东部地方卫生区的脆弱人群和优先人群：澳大利亚统计局2016年人口普查分析。https://www.seslhd.health.nsw.gov.au/sites/default/files/groups/PICH/Priority%20Populati ons/Population%20Profile_vulnerable%20and%20priority%20populations_SESLHD_ABS%20 Census%202016.pdf

46. 伊拉瓦拉肖尔黑文地方卫生区。医疗服务计划2020—2030年。https://www.islhd.health.nsw.gov.au/sites/default/files/Health%20Plans/Final%20Health%2 0Care%20Services%20Plan%202020-2030.pdf 47. Nghiem S, Afoakwah C, Scuffham P, Byrnes J. 昆士兰心脏病学记录关联队列（QCard）研究的基线概况。《BMC心血管疾病》2022年;22. doi: 10.1186/s12872-022-02478-z 48. Indraratna P, Biswas U, McVeigh J, et al. TeleClinical Care：针对心力衰竭或急性冠脉综合征患者的基于智能手机的照护模式随机对照试验。《心脏、肺与循环》2021年;30:S101. doi: https://doi.org/10.1016/j.hlc.2021.06.026 49. Wei W-Q, Denny JC. 从电子健康记录中提取研究级表型以支持精准医学。《基因组医学》2015年;7. doi: 10.1186/s13073-015- 0166-y 50. Weiskopf NG, Weng C. 电子健康记录数据质量评估的方法和维度：促进其在临床研究中的再利用。《美国医学信息学会杂志》2013年;20:144-151. doi: 10.1136/amiajnl-2011-000681 51. 观察性健康数据科学与信息学（OHDSI）。OMOP通用数据模型。https://www.ohdsi.org/data-standardization/the-common-data-model/（2021年12月15日 2021年）52. Quiroz JC, Chard T, Sa Z, et al. 将健康数据库转换为OMOP的提取、转换、加载框架。《公共科学图书馆·综合》2022年;17:e0266911. doi: 10.1371/journal.pone.0266911 53. Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR. 通过评估观察性数据质量，提升对真实世界证据的信任。《美国医学信息学会杂志》2021年;28:2251-2257. doi: 10.1093/jamia/ocab132 54. 观察性健康数据科学与信息学。观察性健康数据科学与信息学。https://www.ohdsi.org/（30/03/2023）55. Khare R, Utidjian LH, Razzaghi H, et al. 大型儿科研究网络数据质量评估工作流程的设计与完善。EGEMS（华盛顿特区）2019年;7:36. doi: 10.5334/egems.294

补充材料

电子研究机构云架构（ERICA）

电子研究机构云架构（ERICA）平台由新南威尔士大学建立。ERICA 为使用敏感微观数据开展研究提供安全云（亚马逊云科技，AWS）计算环境。ERICA 的设置和配置高度自动化，最大限度地降低了人为错误的可能性。在所有外部访问点，用户均使用一组登录凭据（账户

名和密码）以及强制性的多因素身份验证代码（通过智能手机）进行身份验证。研究人员通过远程访问 Windows 虚拟工作站使用 ERICA 环境，这些工作站配备防病毒软件和恶意软件检测系统。研究人员无法从其 ERICA 项目工作区访问任何外部网站、电子邮件或其他通信渠道。ERICA 中保存的所有研究数据在静态存储和传输过程中均经过加密，所有通信和数据传输均使用安全协议。研究人员只能通过称为 Hub 的受控且经过审计的网关机制导入或导出数据（外部 eHub 和内部 iHub 应用程序），所有数据传输均会被完整记录，并接受病毒检查和完整副本审计跟踪。指定的项目控制员负责检查并批准所有入站和出站文件。包括剪贴板、电子邮件、即时通信服务、打印服务和互联网访问在内的所有其他文件或数据进入及导出机制均被阻止。项目工作区彼此隔离，除非通过 Hub，否则无法在工作区之间传输数据。所有用户都必须接受有关伦理、隐私、信息安全、统计披露控制以及 ERICA 使用方面的培训。数据会定期进行现场和异地备份。所有异地备份和归档数据在传输至安全的异地存储设施之前均经过加密。ERICA 已完成 eHealth NSW 隐私与安全保障框架（PSAF）的全面审查，eHealth NSW 已正式告知，ERICA 的设计具备充分的安全控制措施，可保护新南威尔士州卫生部的数据。

有关 ERICA 的更多信息，请访问 https://research.unsw.edu.au/erica

CardiacAI 数据说明

SESLHD 和 ISLHD 之前属于同一个地方卫生区，因此共用一个电子病历实例。因此，收治于威尔士亲王医院、圣乔治医院、The Sutherland Hospital的患者

或 The Wollongong Hospital 在所有院区均使用同一个患者ID。电子病历住院病例（就诊记录）附有就诊ID，可唯一标识每次住院收治。个人身份标识符不会转移至 CardiacAI 数据存储库。患者ID 和就诊ID 是源电子病历数据库后端的唯一标识符，无法在前端电子病历应用程序中搜索。仅就诊于急诊科且未被收治为住院患者的患者目前不包含在 CardiacAI 数据存储库中。

从电子病历中提取的数据包括结构化数据和非结构化数据类型。电子病历数据元素列于 表6。完整的资料库数据字典可在以下网址找到：www.CardiacAI.org/data-dictionary。病理学和生命体征测量值是根据专家临床意见，结合其与项目目标的相关性进行选择的。完整的纳入测量指标列表可在以下网址找到：www.CardiacAI.org/measurement-concepts。电子病历数据包括患者入院期间创建的完整非结构化临床文档列表。资料库中包含的完整文档类型列表可在以下网址找到：www.CardiacAI.org/document-concepts。

表6. CardiacAI 数据类型

数据类型	**数据元素**
结构化	患者人口统计信息 医院就诊详情 病房转移

	分诊类别 入院诊断代码（SNOMED） 诊断及其他病症 过敏 手术病例详情* 手术操作代码（ACHI）* 药物医嘱 药物给药 入院用药史 出院药物医嘱 用药核对事件 血液制品输注 病理检查结果 微生物学检查结果 医学影像检查详情** 生命体征

非结构化数据	分诊报告 入院前表格 手术报告* 临床病程记录 临床恶化/护理升级记录 心理健康评估 出院小结 协同医疗记录 评估与护理计划 心电图记录（目前不包含影像和信号数据）
* 手术病例数据不包括心脏导管插入术。 ** 医学影像数据不包含实际图像。	

人群健康数据集数据字典

可通过以下途径查找关联人口健康数据的数据字典。

链接:

表7. 人群健康数据集数据字典

人群健康数据集	数据字典链接
住院患者数据收集 (APDC)	https://www.cherel.org.au/media/38875/admitted-patient-data-collection_data-词典_aug_2022.docx
急诊科数据收集 (EDDC)	https://www.cherel.org.au/media/60057/emergency-department-data-collection-data-dictionary_aug_2022.docx
出生、死亡登记处 婚姻（RBDM）与死因 死亡记录单元文件（COD-URF）	https://www.cherel.org.au/media/60080/新南威尔士州-mortality-数据字典_三月_2020.docx

最常见的其他诊断和手术代码

CardiacAI 数据存储库中非急性冠脉综合征、心力衰竭、心房颤动或晕厥住院病例的主要诊断和操作代码。

表 8 其他就诊类型中排名前五的主要诊断和操作代码

		result
Primary Diagnosis	R07.4 - Chest pain, unspecified	3781 (13.99%)
	R07.3 - Other chest pain	1081 (4.0%)
	I70.21 - Atherosclerosis of arteries of extremities with intermittent claudication	814 (3.01%)
	Z45.0 - Adjustment and management of cardiac device	797 (2.95%)
	I35.0 - Aortic (valve) stenosis	699 (2.59%)
Primary Procedure	95550-03 - Allied health intervention, physiotherapy	1324 (7.0%)
	35303-06 - Percutaneous transluminal balloon angioplasty	1166 (6.17%)
	95550-09 - Allied health intervention, pharmacy	1140 (6.03%)
	38353-00 - Insertion of subcutaneous cardiac pacemaker generator	1095 (5.79%)
	38218-00 - Coronary angiography with left heart catheterisation	1077 (5.7%)