

基于医疗保险数据库的皮肤型红斑狼疮及其亚型快速识别规则构建与应用



王予童^{1, 2}, 孟祥龙^{1, 2}, 潘雨^{3, 4}, 尉晨⁵, 靳慧^{3, 4}, 王胜锋^{1, 2, 6}

1. 北京大学公共卫生学院流行病与卫生统计学系 (北京 100191)
2. 教育部重大疾病流行病学重点实验室 (北京 100191)
3. 中国医学科学院、北京协和医学院皮肤病医院/中国医学科学院、北京协和医学院皮肤病研究所 (南京 210042)
4. 中国医学科学院免疫性皮肤病基础与转化研究重点实验室 (南京 210042)
5. 上海崧胜商务咨询有限公司 (上海 201913)
6. 北京大学人工智能研究院 (北京 100871)

【摘要】目的 构建并验证皮肤型红斑狼疮 (CLE) (包含未明确分型的患者) 及其 2 种亚型 [盘状红斑狼疮 (DLE) 和亚急性皮肤型红斑狼疮 (SCLE)] 的数据提取与患者识别规则, 以实现从医疗保险数据库中高效识别 CLE 患者。**方法** 利用 2013—2017 年全国医疗保险数据库, 构建数据提取与患者识别规则, 并以人工核查结果为金标准, 计算规则的灵敏度和特异度, 同时对识别出的患者进行基本特征分析。**结果** 基于标准医学术语及诊断编码构建标准表达式, 并结合临床医师的经验, 补充可能的别称、错写等情况, 以完善初筛表达式。在此基础上, 由临床医师与数据管理工程师通过反复核查, 最终确定疾病精筛规则。构建的 3 种目标疾病的数据提取与患者快速识别规则均表现良好, CLE、DLE 和 SCLE 的灵敏度分别为 0.985、1.000 和 0.991, 特异度分别为 0.997、0.999 和 0.998。共提取 2013—2017 年 CLE 患者 34 554 例, 其中 DLE 患者 2 879 例, SCLE 患者 623 例, 女性比例均高于男性。**结论** 构建的 CLE 数据提取与患者识别规则具备良好的性能, 为皮肤病学研究中患者快速识别方法的开发与优化提供了参考。

【关键词】 皮肤型红斑狼疮; 医疗保险数据; 数据快速提取规则

【中图分类号】 R 593.24 **【文献标识码】** A

Development and application of a rapid identification algorithm for cutaneous lupus erythematosus and its subtypes based on medical insurance databases

WANG Yutong^{1,2}, MENG Xianglong^{1,2}, PAN Yu^{3,4}, WEI Chen⁵, JIN Hui^{3,4}, WANG Shengfeng^{1,2,6}

1. Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China
2. Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China
3. Hospital for Skin Diseases, Institute of Dermatology, Chinese Academy of Medical Sciences and Peking Union Medical College, Nanjing 210042, China

DOI: 10.12173/j.issn.1005-0698.202504078

基金项目: 国家自然科学基金专项项目 (72342015)

通信作者: 靳慧, 博士, 主治医师, Email: jinhui333@pumcdm.cams.cn

王胜锋, 博士, 研究员, 博士研究生导师, Email: wangshengfeng@bjmu.edu.cn

<https://ywlbxb.whuzhmedj.com/>

4. Key Laboratory of Basic and Translational Research for Immune-Mediated Skin Diseases, Chinese Academy of Medical Sciences, Nanjing 210042, China

5. Shanghai Songsheng Business Consulting Co. Ltd., Shanghai 201913, China

6. Institute for Artificial Intelligence, Peking University, Beijing 100871, China

Corresponding authors: JIN Hui, Email: jinhui333@pumcdern.cams.cn; WANG Shengfeng, Email: wangshengfeng@bjmu.edu.cn

【Abstract】Objective To develop and validate data extraction and patient identification algorithms for cutaneous lupus erythematosus (CLE) and its two subtypes, discoid lupus erythematosus (DLE) and subacute cutaneous lupus erythematosus (SCLE), and to enable high-efficiency patient identification in large-scale electronic health databases. **Methods** This study utilized data from the 2013-2017 National Insurance Claims for Epidemiological Research (NICER) to construct data extraction and rapid patient identification algorithms. The manual verification results were used as gold standard to assess the sensitivity and specificity of the algorithms. Additionally, the basic characteristics of the identified patients were analyzed. **Results** Initially, standardized expressions were developed based on medical terminology and diagnostic codes. These were further refined with input from clinicians to include potential synonyms and common misspellings, improving the preliminary screening expressions. Through iterative verification by clinicians and data management engineers, a final disease-specific screening algorithm was established. The developed extraction and identification algorithms for all 3 targeted disease demonstrated strong performance, with sensitivity values of 0.985, 1.000, and 0.991, and specificity values of 0.997, 0.999, and 0.998 for CLE, DLE, and SCLE, respectively. A total of 34,554 CLE cases, including 2,879 DLE cases, and 623 SCLE cases were identified between 2013 and 2017, with a higher prevalence among females than males. **Conclusion** This study developed and validated an identification algorithm for CLE patients based on medical insurance databases, demonstrating high performance. The proposed algorithm provides a methodological framework and empirical evidence for designing and optimizing big data-driven rapid patient identification algorithms in dermatology research.

【Keywords】 Cutaneous lupus erythematosus; Medical insurance; Rapid identification algorithm

红斑狼疮 (lupus erythematosus, LE) 是一种慢性、反复发作的自身免疫性疾病, 临床表现复杂多样。该病为一种病谱性疾病, 其一端是主要累及皮肤黏膜的皮肤型红斑狼疮 (cutaneous lupus erythematosus, CLE), 另一端为累及多脏器、多系统的系统性红斑狼疮 (systemic lupus erythematosus, SLE) [1]。CLE 与 SLE 之间并无明确界限。研究 [2] 显示, 超过 75% 的 SLE 患者会出现皮肤受累, 其中部分以皮肤症状为首发症状。此外, 部分 CLE 患者在病程中可能进展为 SLE, 成人和儿童患者的进展比例分别为 0%~42% 和 0%~31%。急性皮肤型红斑狼疮 (acute cutaneous lupus erythematosus, ACLE) 被认为是 SLE 的最典型的皮肤表现之一 [3-5]。

根据临床表现和组织病理学特征, CLE 主要分为 ACLE、亚急性皮肤型红斑狼疮 (subacute

cutaneous lupus erythematosus, SCLE) 以及慢性皮肤型红斑狼疮 (chronic cutaneous lupus erythematosus, CCLE) [6]。其中, ACLE 可进一步分为局限型和泛发型; SCLE 可分为环状红斑型和丘疹鳞屑型; CCLE 则涵盖多种亚型, 最常见的是盘状红斑狼疮 (discoid lupus erythematosus, DLE), 可分为局限性和播散性。此外, CCLE 还包括: 疣状 LE、肿胀性 LE、深在性 LE/ 狼疮性脂膜炎、冻疮样 LE、Blaschko 线状 LE [1]。在所有 CLE 亚型中, DLE 和 SCLE 最为多见, 分别占 CLE 总病例的 50%~85% 和 7%~27% [7]。由于病程长, 且直接影响患者外貌, 部分亚型还会遗留瘢痕、脂肪萎缩等损容性后遗症, CLE 不仅严重影响患者生活质量, 还会带来沉重的经济负担 [8]。因此, 了解 CLE 及其常见亚型的流行病学特征, 对于全面掌握疾病现状、实现疾病及时诊断和治

疗具有重要意义^[9]。然而,目前我国尚缺乏针对 CLE 及其常见亚型的大规模基于人群的流行病学特征研究。

医疗保险数据是在医疗保险业务过程中产生的数据,具有数据量庞大、覆盖范围广的特点,可较为完整地记录特定时间内人群的就诊信息和报销情况,为患者治疗模式、医疗资源使用及成本探索提供真实世界证据^[10]。但既往研究发现,医疗保险数据主要用于医疗报销,并非专为研究而设,简单使用疾病编码可能无法直接获取研究所需的病例。因此,在利用这些数据开展流行病学研究(如疾病负担估算)或临床研究时,需根据研究目标制定合适的疾病定义^[11]。基于此,本研究利用 2013—2017 年全国医疗保险数据库,构建适用于 CLE 及其 2 种常见亚型(DLE 和 SCLE)的数据提取与患者识别规则,以期为基于健康医疗大数据估算疾病负担的研究提供参考。

1 资料与方法

1.1 数据来源

研究数据来源于 2013—2017 年全国医疗保险流行病学研究数据库,该数据涵盖了城镇职工基本医疗保险(urban employee basic medical insurance, UEBMI)和城镇居民基本医疗保险(urban resident basic medical insurance, URBMI)数据^[12-13],截至 2016 年已覆盖中国超过 95% 的城镇人口^[14-15]。其中,UEBMI 数据主要覆盖城市地区在职和退休员工,包括政府机构、事业单位、国有企业、私营企业、社会组织和其他私营实体的雇主和雇员;URBMI 数据涵盖城市地区中未参加职工医疗保险的居民,即儿童、学生、老年人和失业居民。研究纳入的变量包括年龄、性别、就诊年份、医疗保险类型。所有数据对个人信息均进行匿名化处理。本研究经北京大学生物伦理委员会审批通过(批件号:IRB00001052-24068),并豁免知情同意。

1.2 数据提取和患者识别规则的构建

研究数据提取的主要框架与既往研究一致^[13],具体为:首先,筛选包含“狼疮”或“LE”的诊断以及国际疾病分类第九版(International Classification of Diseases-9, ICD-9)与第十版(ICD-10)编码,构建 LE 数据库(表 1),具体检索式可见 ResearchGate DOI: [10.13140/RG.2.2.23330.34242](https://doi.org/10.13140/RG.2.2.23330.34242)。

在 LE 数据库基础上,利用表达式进一步提取 CLE 数据库。CLE 的病例定义为患者在医疗保险数据库中曾被诊断为任一 CLE 亚型(如 DLE、SCLE、ACLE 等)或有 CLE 相关诊断但亚型分型尚不明确,如曾被诊断为红斑狼疮(L93.001)、局限性红斑狼疮(其他的)(L93.200),无论其是否合并存在 SLE 诊断记录^[16]。疾病表达式主要分为两类:标准诊断表达式(set ①)和其他相关表达式(set ②)。其中,标准诊断表达式为基于 ICD-9、ICD-10 编码以及国家卫生健康委员会于 2016 年发布的国家标准《GB/T 14396-2016 疾病分类与代码》中的中文医学术语^[17]。此外,在真实临床环境中,缩写、别名及拼写错误较为常见,使得疾病表述可能有所变化^[13]。因此,本研究进一步构建了可能与疾病相关的其他相关表达式,以最大程度涵盖临床记录中的多样性。该部分由皮肤科专科医师依据临床经验独立制定,尽可能列举了各种缩写、别名及潜在拼写错误,以提高筛选的全面性和准确性。简而言之,其他相关表达式不仅涵盖了 CLE,还可能涉及其他具有狼疮样表现的疾病。相较于标准诊断表达式,这一表达式组合能更广泛地捕捉潜在 CLE 病例,但其筛选结果可能会遗漏部分 CLE 病例或将部分非 CLE 病例误归类在其中,仍需进一步判定。

在 LE 数据库提取到的 78 585 条诊断文本中,5 614 条诊断文本包含 CLE 标准诊断的术语或编码。针对剩余诊断文本,由临床医师依据纳入与排除标准进行人工核查,并提取符合要求的诊断文本及其对应编码。具体纳入标准:① LE 及其各亚型(包括各型 CLE 和 SLE)的标准诊断及 ICD-10、ICD-9 编码;② LE 及其各亚型相关诊断别称、可能的错误拼写名称;③符合 LE 及其各亚型特异性相关临床表现描述(如蝶形红斑、盘状红斑)^[1, 18]。排除标准:①诊断文本无法证明为诊断明确的 LE/CLE/DLE/SCLE;②疾病名称中含有相关筛选字符,但明确不是目标疾病患者,如“寻常狼疮”(为皮肤结核的一种);③因检索词局限型带来的错误筛选患者,如因“LE”检索词带来的“利普刀(LEEP)术后”患者;④疾病待查、疑诊(例如诊断中包含“?”等符号)或不确定诊断(LE 集合为总集未排除, CLE/DLE/SCLE 筛选集均有排除)。随后,与数

表1 LE数据库的具体筛选策略
Table 1. Selection strategy for the LE database

诊断信息类别	目标检索词
标准诊断及 其ICD-10或 ICD-9编码	ICD-10编码体系：红斑狼疮（L93.001）、盘状红斑狼疮（L93.000）、盘状狼疮（L93.000x002）、局限性红斑狼疮（其他的）（L93.200）、亚急性皮肤红斑狼疮（L93.100）、急性皮肤红斑狼疮（L93.200x001）、肿胀性（瘤样）狼疮（L93.200x003）、狼疮性脂膜炎（L93.201）、深在性红斑狼疮（L93.202）、药物性系统性红斑狼疮（M32.000）、系统性红斑狼疮、累及器官或系统（M32.100）、狼疮抗凝物质出现（D68.600x011）、中毒性肝病伴狼疮样肝炎（K71.500x002）、狼疮样肝炎（K73.200x001）、狼疮性肝炎（K75.400x001）、狼疮性肝损害（M32.108+K77.8*）、狼疮性肾损害（M32.100x001）、狼疮性肾炎（M32.101+、M32.101+N08.5*）、狼疮性肾小管间质肾炎（M32.102+N16.4*）、狼疮性肺病变（M32.100x007）、狼疮性肺炎（M32.103+J99.1*）、系统性红斑狼疮所致精神障碍（F06.800x021）、狼疮性痴呆（M32.100x021）、狼疮性神经炎（M32.100x008）、狼疮性周围神经病（M32.106+G63.5*）、狼疮性脊髓病变（M32.107+G99.2*）、狼疮性脑病（M32.114+G94.8*）、系统性红斑狼疮合并肌病（M32.100x014+G73.7*）、狼疮性肌病（M32.110+G73.7*）、狼疮性关节炎（M32.100x016）、狼疮累及胃肠道（M32.100x018+K93.8*）、狼疮性胃肠道损害（M32.112+K93.8*）、非典型疣状心内膜炎[Libman-Sacks病]（M32.100x006+I39.8*）、利布曼-萨克斯病（M32.109+I39.8*）、狼疮性心肌病（M32.104+I43.8*）、狼疮性心包炎（M32.105+I32.8*）、狼疮性血液系统损害（M32.111+D77*）、狼疮性视网膜病变（M32.113+H36.8*）、狼疮性浆膜炎（M32.115+K67.8*）、系统性红斑狼疮、其他形式的（M32.800）、系统性红斑狼疮（M32.900）、隐匿性系统性红斑狼疮（M32.901）、妊娠合并系统性红斑狼疮（O99.811）、新生儿狼疮综合征（P00.800x001）、新生儿红斑狼疮（P00.814）、母体系统性红斑狼疮新生儿（P00.807） ICD-9编码体系：红斑狼疮（695.401）、盘状红斑狼疮（695.402）、系统性红斑狼疮（710.001）、狼疮性脑病（710.002+348.304*）、狼疮性心包炎（710.003+420.003*）、狼疮性心肌病（710.004+425.801*）、狼疮性肾炎（710.005+583.806*）、狼疮性肺炎（710.007+517.801*）、狼疮性肺间质纤维化（710.008+517.803*）、狼疮性肌病（710.009+359.603*）
别称/可能的 错误拼写（中 文、英文或中 英文混写）	皮肤狼疮、CLE、cutaneous LE、ACLE、SCLE、慢性皮肤型红斑狼疮、CCLE、DLE、泛发型（播散性）DLE、Disseminated DLE、DDLE、LEP、脂膜炎性狼疮、脂膜炎样狼疮、狼疮脂膜炎、深在狼疮、深在LE、疣状（样）狼疮、疣状LE、VLE、肥厚狼疮、肥厚LE、冻疮样狼疮、CHLE、冻疮LE、冻疮狼疮、肿胀狼疮、LET、间歇狼疮、ICLE、线状（样）狼疮、线状红斑狼疮、线状LE、大疱性（样）狼疮、水疱狼疮、BLE、系统狼疮、systemic LE、妊娠狼疮、妊娠性LE、狼疮肾、狼疮脑、狼疮肝、妊娠狼疮、新生儿狼疮、新生儿LE、婴儿狼疮、婴儿LE、胎儿狼疮、胎儿LE
诊断相关临床 表现描述	蝶形（状）红斑、红蝴蝶、银屑病样狼疮、环状红斑型（样）狼疮、丘疹鳞屑型（样）狼疮
排除诊断	ICD-10编码体系：寻常性狼疮（A18.410）、播散性粟粒性狼疮（A18.400x001）、结核性狼疮（A18.409）、冻疮样狼疮型皮肤结节病（D86.300x002）、眼睑盘形红斑狼疮性皮炎（H01.100x006）、狼疮样须疮（L73.801） ICD-9编码体系：寻常性狼疮（017.003） 排除诊断的不规则命名：寻常狼疮、狼疮结核、粟粒狼疮 因检索词局限性带来的错误筛选诊断汇总：利普刀（LEEP）术后、多囊卵巢综合征（Stein-Leventhal）、癫痫（epilepsy）、Lennox-Gastaut综合征、惠普尔病（Whipple's）、Leri-Weill综合征、Colles骨折、Leigh综合征、腹会阴联合直肠癌根治术（Miles手术）、胰十二指肠根治性切除术（Whipple手术）、IgA肾病Lee氏、IgA血管炎（Henoch-Schönlein purpura）、前列腺癌Gleason分级、Castleman病、Lennox-Gastaut综合征、Liddle综合征、Leopard综合征、Stein-Leventhal综合征、Miller-Fisher综合征、Eales病、Angle分类、类风湿关节炎（带有拼音）、Lewis-Sumner综合征、食管肌切开术（改良Heller手术）、布尔纳维氏病（Bourneville病）、Galeazzi骨折、Kellgren-Lawrence分级、Judet-Letournel分型、过敏性接触性皮炎（allergic contact dermatitis）、双侧下肢轻瘫/无力（bilateral LE paresis/weakness）、出血（bleed）、Clerambault-Kandinsky综合征、胆总管占位（common bile duct mass）、寻常疣（common wart）、结节（lesion）、Crigler-Najjar综合征、Ehlers-Danlos综合征、持久性隆起性红斑（erythema elevatum diutinum）、女性型脱发（female-pattern alopecia）、白血病（leukemia）、leber病、Lesch-Nyhan综合征、Leriche综合征、头皮感染/病变（scalp infection/lesion）、Gitelman综合征、Glenn手术、Henoch-Schönlein紫癜、单纯疱疹、K-F环（Kayser-Fleischer ring）、little、leg、left、脊柱侧弯Lenke分型、系统性硬化症、Wallenberg综合征、Dressler综合征、骨盆骨折（Tile分型）、淋巴水肿矫正术（Charles手术）、虹膜角膜内皮综合征、小肠折叠术（Noble手术）、腹膜后苗勒管残余囊肿、Wheeler睑缘缝合术、Müller-Weiss病、子宫恶性Müllerian混合瘤（癌肉瘤）

据工程师进行反复核对，排除所有待查或不确定诊断，直至所有诊断文本均符合疾病诊断标准，最终确定用于进一步筛选的其他表达式(set ③)。该表达式(set ③)与前述标准诊断表达式(set ①)合并，共同构成最终的精准筛选表达式。为验证所构建的数据提取表达式的有效性和准确性，进一步对原始数据库中的每条诊断文本进行人工判定（其中，CLE 以 LE 数据库为基础筛选，DLE 和 SCLE 以 CLE 数据库为基础筛选），判断其是否属于目标疾病。最终，以人工核查结果

作为“金标准”，计算表达式的灵敏度和特异度^[13, 19]。根据前期研究^[14]，仅保留数据质量较高的 23 个省份的数据。完整流程见图 1。在通过表达式提取所有符合标准的诊断记录后，使用患者唯一标识符对不同年份信息进行整合与去重，仅保留首次确诊的 CLE 亚型诊断记录^[20]，用于后续的发病特征分析。在对抓取的 CLE 患者及诊断为 DLE 与 SCLE 2 种亚型患者的基本特征进行描述中，计量资料以 $\bar{x} \pm s$ 表示，计数资料以 $n(%)$ 表示。

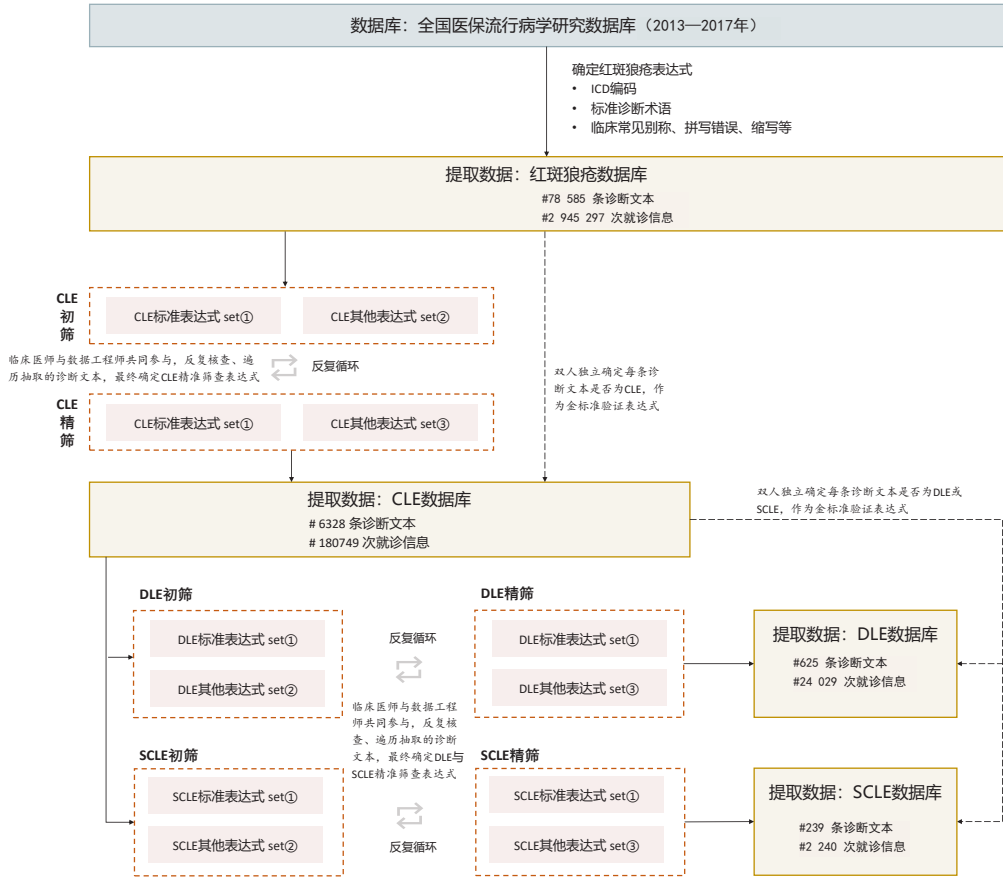


图1 构建CLE及其2种亚型 (DLE以及SCLE) 数据提取和患者识别规则的流程

Figure 1. Data extraction and patient identification process for CLE and its subtypes (DLE and SCLE)

注：CLE. 皮肤型红斑狼疮；DLE. 盘状红斑狼疮；SCLE. 亚急性皮肤型红斑狼疮。

2 结果

最终确定的数据提取与患者识别规则涵盖以下 3 个方面：①标准诊断名称及其对应的 ICD 编码（包括 ICD-9 和 ICD-10）；②疾病的别称或可能存在的拼写错误（包括中文、英文以及中英文混写形式）；③与诊断相关的临床表现描述。在 CLE 中，标准诊断名称包括红斑

狼疮、盘状红斑狼疮、局限性红斑狼疮（其他的）、亚急性皮肤红斑狼疮、急性皮肤红斑狼疮、肿胀性（瘤样）狼疮、狼疮性脂膜炎、深在性红斑狼疮；相应 ICD 编码为 ICD-10 的 L93 和 ICD-9 的 695.4。在别称及拼写错误方面，除了常见英文缩写外，还包括银屑病狼疮、间歇狼疮、冻疮狼疮、疣样（状）LE、肥厚狼疮、线样（状）狼疮、大疱狼疮、水疱狼疮等拼写不规范情况；

在诊断相关临床表现描述中包含了环形（状）红斑狼疮、丘疹鳞屑狼疮、蝶形（状）红斑及“红蝴蝶”等描述词。最后，为保证数据准确性，进一步排除疑似或待查诊断条目，避免因诊断不明影响结果准确性。

在 CLE 数据库中，进一步提取 DLE 以及 SCLE 相关诊断，包括标准诊断名称、对应 ICD 编码以及别称或可能的错误拼写。具体而言，在 DLE 中，标准诊断为盘状红斑狼疮；ICD 编码包括 L93.0（ICD-10）以及 695.402（ICD-9）；别称以及可能的错误拼写为 DLE。在 SCLE 中，标准诊断为亚急性皮肤红斑狼疮；ICD 编码为 L93.1（ICD-10）；别称以及拼写错误包括 SCLE、环

状（形）红斑以及丘疹鳞屑（表 2）。基于最终确定的诊断文本，提取关键诊断特征，用于构建系统性的数据提取与患者识别规则检索策略以人工判定结果被作为金标准，CLE、DLE 以及 SCLE 三种目标疾病的数据提取和患者识别规则表现均良好，灵敏度分别为 0.985、1.000 以及 0.991，特异度分别为 0.997、0.999 以及 0.998。

在医疗保险数据库中共提取 2013—2017 年 CLE 患者 34 554 例，平均年龄为（44.21 ± 15.05）岁，女性 26 904 例（77.86%）。DLE 患者和 SCLE 患者分别有 2 879 例以及 623 例，诊断时的平均年龄分别为（48.10 ± 15.32）岁和（42.96 ± 15.74）岁，女性占比均高于男性（表 3）。

表2 在LE数据库中的CLE、DLE以及SCLE的数据提取和患者识别规则
Table 2. Data extraction and patient identification of CLE, DLE, and SCLE cases from the lupus erythematosus database

目标疾病	目标检索词
CLE	
标准诊断及 ICD-10或ICD-9 编码	ICD-10编码体系：盘状红斑狼疮（L93.000）、盘状狼疮（L93.000x002）、局限性红斑狼疮（其他的）（L93.200）、亚急性皮肤红斑狼疮（L93.100）、急性皮肤红斑狼疮（L93.200x001）、肿胀性（瘤样）狼疮（L93.200x003）、狼疮性脂膜炎（L93.201）、深在性红斑狼疮（L93.202） ICD-9编码体系：红斑狼疮（695.401）、盘状红斑狼疮（695.402）
别称/可能的错误 拼写（中文、 英文或中英文 混写）	皮肤狼疮、CLE、cutaneous LE、ACLE、SCLE、慢性皮肤型红斑狼疮、CCLE、DLE、泛发型（播散性）DLE、Disseminated DLE、DDLE、LEP、脂膜炎性狼疮、脂膜炎样狼疮、狼疮脂膜炎、深在狼疮、深在LE、疣状（样）狼疮、疣状LE、VLE、肥厚狼疮、肥厚LE、冻疮样狼疮、CHLE、冻疮LE、冻疮狼疮、肿胀狼疮、LET、间歇狼疮、ICLE、线状（样）狼疮、线状红斑狼疮、线状LE、大疱性（样）狼疮、水疱狼疮、BLE
诊断相关临床 表现描述	蝶形（状）红斑、红蝴蝶、环状红斑型（样）狼疮、丘疹鳞屑型（样）狼疮、银屑病样狼疮
排除诊断	因检索词局限性带来的错误筛选诊断汇总：慢性病狼疮 排除疑似诊断和待查诊断：“狼疮？”“红蝴蝶？”“皮疹待查盘状狼疮？”“脂膜炎？”
DLE	
标准诊断及 ICD-10或ICD-9编码	ICD-10编码体系：盘状红斑狼疮（L93.000）、盘状狼疮（L93.000x002） ICD-9编码体系：盘状红斑狼疮（695.402）
别称/可能的错误 拼写（中文、英文 或中英文混写）	泛发型（播散性）DLE、disseminated DLE、DDLE
SCLE	
标准诊断及 ICD-10或ICD-9编码	ICD-10编码体系：亚急性皮肤红斑狼疮（L93.100） 环状红斑型（样）狼疮、丘疹鳞屑型（样）狼疮、银屑病样狼疮、SCLE
别称/可能的错误 拼写（中文、英文 或中英文混写）	

注：LE. 红斑狼疮；CLE. 皮肤型红斑狼疮；DLE. 盘状红斑狼疮；SCLE. 亚急性皮肤型红斑狼疮。

表3 提取的CLE及其2种亚型患者基本特征[n, $\bar{x} \pm s$, n (%)]
Table 3. Basic characteristics of patients with CLE and its two subtypes [n, $\bar{x} \pm s$, n (%)]

特征	CLE			DLE			SCLE		
	总人群	女性	男性	总人群	女性	男性	总人群	女性	男性
例数	345 54	26 904	7 650	2 879	1 981	898	623	470	153
年龄 (岁)	44.21 ± 15.05	43.00 ± 14.61	48.50 ± 15.80	48.10 ± 15.32	46.93 ± 15.38	50.68 ± 14.88	42.96 ± 15.74	41.14 ± 15.20	48.56 ± 16.09
就诊年份									
2013年	8 406 (24.33)	7 260 (26.98)	1 146 (14.98)	884 (30.71)	704 (35.54)	180 (20.04)	112 (17.98)	89 (18.94)	23 (15.03)
2014年	7 158 (20.72)	5 881 (21.86)	1 277 (16.69)	572 (19.87)	386 (19.49)	186 (20.71)	132 (21.19)	103 (21.91)	29 (18.95)
2015年	5 029 (14.55)	4 073 (15.14)	956 (12.50)	505 (17.54)	328 (16.56)	177 (19.71)	107 (17.17)	82 (17.45)	25 (16.34)
2016年	7 375 (21.34)	5 179 (19.25)	2 196 (28.71)	419 (14.55)	267 (13.48)	152 (16.93)	121 (19.42)	87 (18.51)	34 (22.22)
2017年	6 586 (19.06)	4 511 (16.77)	2 075 (27.12)	499 (17.33)	296 (14.94)	203 (22.61)	151 (24.24)	109 (23.19)	42 (27.45)
医疗保险类型									
UEBMI	29 604 (85.67)	22 734 (84.50)	6 870 (89.80)	2 457 (85.34)	1 687 (85.16)	770 (85.75)	500 (80.26)	372 (79.15)	128 (83.66)
URBMI	4 950 (14.33)	4 170 (15.50)	780 (10.20)	422 (14.66)	294 (14.84)	128 (14.25)	123 (19.74)	98 (20.85)	25 (16.34)

注：CLE. 皮肤型红斑狼疮；DLE. 盘状红斑狼疮；SCLE. 亚急性皮肤型红斑狼疮；UEBMI. 城镇职工基本医疗保险；URBMI. 城镇居民基本医疗保险。

3 讨论

本研究利用覆盖近 3 亿人的全国医疗保险数据，构建并验证了一种适用于 CLE 及其 2 种亚型（DLE 和 SCLE）的数据提取与患者识别规则，用于识别 CLE、DLE 以及 SCLE 患者。最终规则的表现较好，可以较准确地从医疗保险数据库中识别 CLE 及其亚型病例。本研究共抓取 2013—2017 年 CLE 患者 34 554 例，其中 2 879 例被诊断为 DLE，623 例诊断为 SCLE，为进一步依托全国医疗保险数据测算 CLE 疾病负担及经济负担提供了数据支撑。作为一种慢性、反复迁延的自身免疫性疾病，CLE 目前尚无任何获批的特异性治疗药物^[21]。现有治疗策略主要包括非药物干预（如避免日晒等不良刺激、合理饮食等）、局部治疗（包括糖皮质激素、钙调神经磷酸酶抑制剂及维 A 酸类制剂）以及系统治疗（如抗疟疾药物）^[1]。近年来，针对 CLE 的靶向治疗研发取得进展，包括抑制 I 型干扰素通路的 anifrolumab^[22]、抑制血液树突状细胞抗原 2 的 litifilimab（BIIB059）^[21-23]等新型药物正处于临床试验阶段。了解真实世界中 CLE 患者的总体数量及其流行特征，不仅有助于辅助靶点筛选和优化临床试验设计，也为后续疗效与安全性评估提供重要的数据基础^[24-25]。随着数智化技术的快速发展及电子健康数据的广泛应用，自动化病例识别与数据提取已成为真实世界数据研究的热点方向^[19, 26-27]。既往有研究^[28]总结了构建基于电子健康数据库的病例识别所需要的数据要素，其中使用诊断编码是识别病例最常用与最简单的方法。然而，利用诊断编码识别患者的准确性、有效性与疾病类型以及医生诊断编码的质量密切相关。例如，对于某些疾病（如癌症、糖尿病等），诊断编码的灵敏度和特异度较高，但其他疾病（如心力衰竭）诊断编码的灵敏度和特异度较低^[28-29]。因此，目前的研究尝试整合多种数据要素，以提高病例识别的准确性。

在 LE 疾病领域，现有的快速患者识别规则/算法主要聚焦于 SLE。国内已有研究者同样基于全国医疗保险数据库构建 SLE 患者的识别规则，其灵敏度和特异度均超过 94%，展现出较高的准确性和可靠性^[13]。此外，Barnado 等^[30]基于电子健康记录（electronic health records, EHR）构建了首个识别 SLE 的方法，该方法整合了 ICD-9

编码、实验室指标以及用药记录等多维度临床数据进行患者识别,在训练集和验证集中阳性预测值分别可达到 95% 和 91%。Swerdel 等^[31]则基于医疗保险数据和 EHR 数据,开发了 4 种用于评估 SLE 现患率和新发病例的算法。此外,一些研究者尝试依托医疗保险数据或者医疗保险数据与电子病历数据的整合数据库构建识别 SLE 复发患者的算法^[32-33]。除了上述提到的 SLE 相关研究外,Tang 等^[34]在开发识别哮喘病例的方法时提出,除了就诊记录的诊断外,还可将问题列表(problem list)纳入其中,以捕捉有哮喘记录但是无哮喘就诊诊断的病情较轻的患者,从而更全面地识别目标人群并提高可计算表型的准确性。此外,为进一步优化数据利用,有研究者将通用数据模型与病例识别算法整合,对数据预先进行标准化处理。这种做法不仅提高了数据利用率和识别率,还为向采用不同编码系统的数据库扩展保留了潜力^[26, 35]。

然而,目前尚缺乏针对 CLE 及其常见亚型的快速识别患者的规则/算法的构建,可能与 CLE 皮损异质性较大,导致其准确诊断及分型诊断至今仍是临床工作面临的主要挑战有关^[4]。在已发表的 CLE 疾病负担研究中,仅有一项韩国研究以全国医疗保险数据为资源来源,该研究使用 ICD-10 编码(L93)进行患者识别与病例确定^[36]。然而,一项针对基于医疗保险数据识别 SLE 患者的规则/算法有效性的系统综述^[37]指出,仅依赖 ICD 编码识别患者的方法,在筛选特定人群(如风湿科门诊就诊者)的医疗保险数据时,阳性预测值可达 70%~90%,但在一般人群中仅为 50%~60%。因此,除了标准诊断和疾病编码之外,结合更多的信息,例如更多样的诊断文本表达,可能有助于更全面地识别目标疾病患者并提高识别准确性。以目前 CLE 及其 2 种亚型(DLE 与 SCLE)规则识别出的患者数作为分子,以每年各省各性别年龄亚组城镇职工和城镇居民中参保者对应的总人年作为分母,设置 4 年的洗脱期,估算 2017 年全国 CLE 发病粗率为 9.74/10 万人年,DLE 和 SCLE 的发病率分别为 0.37/10 万人年和 0.14/10 万人年,略高于既往已发表基于人群的 CLE 流行病学研究,可能与现阶段构建的计算表型尚未排除可能合并 SLE 的患者有关,但患者特征与既往研究^[36, 38-39]类似。

目前的 CLE 数据提取与患者识别规则仍存在一定局限,尚有进一步优化的空间。首先,现行疾病分类体系(如 ICD 编码)无法完全对应 CLE 各亚型,这为开发适用于不同 CLE 亚型的患者识别规则带来挑战。已有流行病学研究采用单纯性 CLE(isolated CLE)的定义,即仅纳入具有狼疮特异性皮肤损害但不满足 SLE 分类标准的患者^[40-41],进一步聚焦 CLE 本身的疾病特征,避免 SLE 对疾病负担估计可能的影响。因此,后续研究有必要进一步细化 CLE 的定义,并基于不同亚群评估其疾病负担。其次,本研究中排除了所有待诊待查及疑似诊断的记录。尽管该策略可提升诊断的确定性,但后续研究中可将此类患者纳入分析,并开展敏感性分析以验证识别规则的稳健性^[13]。在外推性方面,当前构建的患者识别规则主要依赖于医疗保险数据库中疾病编码和诊断文本,由于数据覆盖范围广,数据体量大,在国内其他医疗保险数据库具有推广潜力。既往研究^[42-43]表明,医疗保险数据与 EHR 数据在疾病诊断方面具有一定的一致性,然而,不同数据源在诊断记录的完整性和准确性方面仍存在差异^[44],未来可尝试整合 EHR 等多源数据,提升患者识别策略的精度与适用性,实现信息最大化利用^[13]。此外,结合机器学习、大语言模型等方法亦可进一步拓展患者识别规则的覆盖范围并优化其识别效果,未来研究可考虑进一步整合上述技术对现有规则进行优化^[28, 45]。尽管存在上述挑战,本研究基于全国医疗保险数据库,融合标准化的疾病定义与专科医师临床经验,初步建立了可快速识别 CLE 及其常见亚型(DLE 与 SCLE)患者的识别规则,为后续相关研究的开展与优化提供了实证基础与方法参考。

利益冲突声明:作者声明本研究不存在任何经济或非经济利益冲突。

致谢:本研究得到默克雪兰诺(北京)医药研发有限公司的资助,但研究由通信作者发起,默克雪兰诺(北京)医药研发有限公司在研究设计、数据收集与分析、发表决策或论文撰写过程中均未参与。

参考文献

- 1 中华医学会皮肤性病学分会红斑狼疮研究中心. 皮肤型红斑狼疮诊疗指南(2019 版)[J]. 中华皮肤科杂志, 2019, 52(3):

- 149–155. [Centre for Lupus Erythematosus Research, Chinese Society of Dermatology. Guideline for diagnosis and treatment of cutaneous lupus erythematosus(2019)[J]. Chinese Journal of Dermatology, 2019, 52(3): 149–155.] DOI: [10.3760/cma.j.issn.0412-4030.2019.03.001](https://doi.org/10.3760/cma.j.issn.0412-4030.2019.03.001).
- 2 Stull C, Sprow G, Werth VP. Cutaneous involvement in systemic lupus erythematosus: a review for the rheumatologist[J]. J Rheumatol, 2023, 50(1): 27–35. DOI: [10.3899/jrheum.220089](https://doi.org/10.3899/jrheum.220089).
- 3 Curtiss P, Walker AM, Chong BF. A systematic review of the progression of cutaneous lupus to systemic lupus erythematosus[J]. Front Immunol, 2022, 13: 866319. DOI: [10.3389/fimmu.2022.866319](https://doi.org/10.3389/fimmu.2022.866319).
- 4 Lu Q, Long H, Chow S, et al. Guideline for the diagnosis, treatment and long-term management of cutaneous lupus erythematosus[J]. J Autoimmun, 2021, 123: 102707. DOI: [10.1016/j.jaut.2021.102707](https://doi.org/10.1016/j.jaut.2021.102707).
- 5 Gronhagen CM, Foröd CM, Granath F, et al. Cutaneous lupus erythematosus and the association with systemic lupus erythematosus: a population-based cohort of 1088 patients in Sweden[J]. Br J Dermatol, 2011, 164(6): 1335–1341. DOI: [10.1111/j.1365-2133.2011.10272.x](https://doi.org/10.1111/j.1365-2133.2011.10272.x).
- 6 Gilliam JN, Sontheimer RD. Distinctive cutaneous subsets in the spectrum of lupus erythematosus[J]. J Am Acad Dermatol, 1981, 4(4): 471–475. DOI: [10.1016/s0190-9622\(81\)80261-7](https://doi.org/10.1016/s0190-9622(81)80261-7).
- 7 Chang YC, Werth VP. Update on epidemiology and clinical assessment tools of cutaneous lupus erythematosus and dermatomyositis[J]. Curr Dermatol Rep, 2013, 2(1): 48–57. DOI: [10.1007/s13671-012-0037-3](https://doi.org/10.1007/s13671-012-0037-3).
- 8 Tebbe B, Orfanos CE. Epidemiology and socioeconomic impact of skin disease in lupus erythematosus[J]. Lupus, 1997, 6(2): 96–104. DOI: [10.1177/096120339700600204](https://doi.org/10.1177/096120339700600204).
- 9 Fernández Ávila DG, Rivera Maldonado M, Charry Anzola LP. Epidemiology and demographics of cutaneous lupus erythematosus in Colombia between 2015 and 2019[J]. Reumatismo, 2023, 75(1): 23–27. DOI: [10.4081/reumatismo.2023.1568](https://doi.org/10.4081/reumatismo.2023.1568).
- 10 Mathai SC, Hemnes AR, Manaker S, et al. Identifying patients with pulmonary arterial hypertension using administrative claims algorithms[J]. Ann Am Thorac Soc, 2019, 16(7): 797–806. DOI: [10.1513/AnnalsATS.201810-672CME](https://doi.org/10.1513/AnnalsATS.201810-672CME).
- 11 Park JS, Lee CH. Clinical study using healthcare claims database[J]. J Rheum Dis, 2021, 28(3): 119–125. DOI: [10.4078/jrd.2021.28.3.119](https://doi.org/10.4078/jrd.2021.28.3.119).
- 12 Tian Y, Liu H, Si Y, et al. Association between temperature variability and daily hospital admissions for cause-specific cardiovascular disease in urban China: a national time-series study[J]. PLoS Med, 2019, 16(1): e1002738. DOI: [10.1371/journal.pmed.1002738](https://doi.org/10.1371/journal.pmed.1002738).
- 13 Li M, Li C, Cao M, et al. Incidence and prevalence of systemic lupus erythematosus in urban China, 2013–2017: a nationwide population-based study[J]. Sci Bull (Beijing), 2024, 69(19): 3089–3097. DOI: [10.1016/j.scib.2024.04.075](https://doi.org/10.1016/j.scib.2024.04.075).
- 14 Zhang C, Feng J, Wang S, et al. Incidence of and trends in hip fracture among adults in urban China: a nationwide retrospective cohort study[J]. PLoS Med, 2020, 17(8): e1003180. DOI: [10.1371/journal.pmed.1003180](https://doi.org/10.1371/journal.pmed.1003180).
- 15 Li Y, Malik V, Hu FB. Health insurance in China: after declining in the 1990s, coverage rates rebounded to near-universal levels by 2011[J]. Health Aff (Millwood), 2017, 36(8): 1452–1460. DOI: [10.1377/hlthaff.2016.1658](https://doi.org/10.1377/hlthaff.2016.1658).
- 16 Vale ECSD, Garcia LC. Cutaneous lupus erythematosus: a review of etiopathogenic, clinical, diagnostic and therapeutic aspects[J]. An Bras Dermatol, 2023, 98(3): 355–372. DOI: [10.1016/j.abd.2022.09.005](https://doi.org/10.1016/j.abd.2022.09.005).
- 17 国家卫生健康委. 疾病分类与代码 (GB/T 14396–2016) [S/OL]. (2016–10–13) [2025–05–29]. <https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=8127A7785CA677952F9DA062463CBC41>.
- 18 Worm M, Zidane M, Eisert L, et al. S2k guideline: diagnosis and management of cutaneous lupus erythematosus–Part 1: Classification, diagnosis, prevention, activity scores[J]. J Dtsch Dermatol Ges, 2021, 19(8): 1236–1247. DOI: [10.1111/ddg.14492](https://doi.org/10.1111/ddg.14492).
- 19 Yan S, Melnick K, He X, et al. Developing a computable phenotype for glioblastoma[J]. Neuro Oncol, 2024, 26(6): 1163–1170. DOI: [10.1093/neuonc/noad249](https://doi.org/10.1093/neuonc/noad249).
- 20 Tian H, Yang W, Hu Y, et al. Estimating cancer incidence based on claims data from medical insurance systems in two areas lacking cancer registries in China[J]. EclinicalMedicine, 2020, 20: 100312. DOI: [10.1016/j.eclim.2020.100312](https://doi.org/10.1016/j.eclim.2020.100312).
- 21 Cho SK, Vazquez T, Werth VP. Litifilimab (BIIB059), a promising investigational drug for cutaneous lupus erythematosus[J]. Expert Opin Investig Drugs, 2023, 32(5): 345–353. DOI: [10.1080/13543784.2023.2212154](https://doi.org/10.1080/13543784.2023.2212154).
- 22 Xie L, Lopes Almeida Gomes L, Stone CJ, et al. An update on clinical trials for cutaneous lupus erythematosus[J]. J Dermatol, 2024, 51(7): 885–894. DOI: [10.1111/1346-8138.17161](https://doi.org/10.1111/1346-8138.17161).
- 23 Cho YM, Furie R. The development of litifilimab (BIIB 059) for cutaneous and systemic lupus erythematosus[J]. Immunotherapy, 2024, 16(1): 15–20. DOI: [10.2217/imt-2023-0086](https://doi.org/10.2217/imt-2023-0086).
- 24 Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence—What is it and what can it tell us?[J]. N Engl J Med, 2016, 375(23): 2293–2297. DOI: [10.1056/NEJMs1609216](https://doi.org/10.1056/NEJMs1609216).
- 25 Girman CJ, Ritchey ME, Lo Re V 3rd. Real-world data: assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products[J]. Pharmacoepidemiol Drug Saf, 2022, 31(7): 717–720. DOI: [10.1002/pds.5444](https://doi.org/10.1002/pds.5444).
- 26 Wang HY, Du J, Yang Y, et al. Rapid identification of chronic kidney disease in electronic health record database using computable phenotype combining a common data model[J]. Chin Med J (Engl), 2023, 136(7): 874–876. DOI: [10.1097/cm9.0000000000002168](https://doi.org/10.1097/cm9.0000000000002168).
- 27 Dawwas GK, Weiss A, Constant BD, et al. Development and validation of claims-based definitions to identify incident and prevalent inflammatory bowel disease in administrative healthcare

- databases[J]. *Inflamm Bowel Dis*, 2023, 29(12): 1993–1996. DOI: [10.1093/ibd/izad053](https://doi.org/10.1093/ibd/izad053).
- 28 McBrien KA, Souri S, Symonds NE, et al. Identification of validated case definitions for medical conditions used in primary care electronic medical record databases: a systematic review[J]. *J Am Med Inform Assoc*, 2018, 25(11): 1567–1578. DOI: [10.1093/jamia/ocy094](https://doi.org/10.1093/jamia/ocy094).
 - 29 Hirsch AG, Scheck McAlearney A. Measuring diabetes care performance using electronic health record data: the impact of diabetes definitions on performance measure outcomes[J]. *Am J Med Qual*, 2014, 29(4): 292–299. DOI: [10.1177/1062860613500808](https://doi.org/10.1177/1062860613500808).
 - 30 Barnado A, Casey C, Carroll RJ, et al. Developing electronic health record algorithms that accurately identify patients with systemic lupus erythematosus[J]. *Arthritis Care Res (Hoboken)*, 2017, 69(5): 687–693. DOI: [10.1002/acr.22989](https://doi.org/10.1002/acr.22989).
 - 31 Swerdel JN, Ramcharan D, Hardin J. Using a data-driven approach for the development and evaluation of phenotype algorithms for systemic lupus erythematosus[J]. *PLoS One*, 2023, 18(2): e0281929. DOI: [10.1371/journal.pone.0281929](https://doi.org/10.1371/journal.pone.0281929).
 - 32 Goetz I, Choong C, Winnie J, et al. Development of a claims-based flare algorithm for systemic lupus erythematosus[J]. *Curr Med Res Opin*, 2022, 38(9): 1641–1649. DOI: [10.1080/03007995.2022.2101804](https://doi.org/10.1080/03007995.2022.2101804).
 - 33 Wu SS, Perry A, Zimmerman NM, et al. Predictors of flare-related inpatient or emergency department stay in systemic lupus erythematosus: a real-world analysis of Medicaid claims in the United States[J]. *J Manag Care Spec Pharm*, 2024, 30(1): 61–70. DOI: [10.18553/jmcp.2024.30.1.61](https://doi.org/10.18553/jmcp.2024.30.1.61).
 - 34 Tang M, Goldstein BA, He J, et al. Performance of a computable phenotype for pediatric asthma using the problem list[J]. *Ann Allergy Asthma Immunol*, 2020, 125(5): 611–613. e1. DOI: [10.1016/j.anai.2020.07.009](https://doi.org/10.1016/j.anai.2020.07.009).
 - 35 Graham J, Iverson A, Monteiro J, et al. Applying computable phenotypes within a common data model to identify heart failure patients for an implantable cardiac device registry[J]. *Int J Cardiol Heart Vasc*, 2022, 39: 100974. DOI: [10.1016/j.ijcha.2022.100974](https://doi.org/10.1016/j.ijcha.2022.100974).
 - 36 Baek YS, Park SH, Baek J, et al. Cutaneous lupus erythematosus and its association with systemic lupus erythematosus: a nationwide population-based cohort study in Korea[J]. *J Dermatol*, 2020, 47(2): 163–165. DOI: [10.1111/1346-8138.15162](https://doi.org/10.1111/1346-8138.15162).
 - 37 Moores KG, Sathe NA. A systematic review of validated methods for identifying systemic lupus erythematosus (SLE) using administrative or claims data[J]. *Vaccine*, 2013, 31 Suppl 10: K62–73. DOI: [10.1016/j.vaccine.2013.06.104](https://doi.org/10.1016/j.vaccine.2013.06.104).
 - 38 Maillat M, Tounkara BS, Vagner D, et al. Epidemiology of isolated cutaneous lupus erythematosus in the multiethnic population of Reunion Island (France): a retrospective multicentric study[J]. *J Rheumatol*, 2025, 52(3): 251–256. DOI: [10.3899/jrheum.2024-0661](https://doi.org/10.3899/jrheum.2024-0661).
 - 39 Petersen MP, Möller S, Bygum A, et al. Epidemiology of cutaneous lupus erythematosus and the associated risk of systemic lupus erythematosus: a nationwide cohort study in Denmark[J]. *Lupus*, 2018, 27(9): 1424–1430. DOI: [10.1177/0961203318777103](https://doi.org/10.1177/0961203318777103).
 - 40 Durosaro O, Davis MD, Reed KB, et al. Incidence of cutaneous lupus erythematosus, 1965–2005: a population-based study[J]. *Arch Dermatol*, 2009, 145(3): 249–253. DOI: [10.1001/archdermatol.2009.21](https://doi.org/10.1001/archdermatol.2009.21).
 - 41 Maillat M, Tounkara BS, Vagner D, et al. Epidemiology of isolated cutaneous lupus erythematosus in the multiethnic population of Reunion Island: a retrospective multicenter study[J]. *J Rheumatol*, 2025, 52(3): 251–256. DOI: [10.3899/jrheum.2024-0661](https://doi.org/10.3899/jrheum.2024-0661).
 - 42 Lusk JB, Choi S, Clark AG, et al. Dementia and Parkinson's disease diagnoses in electronic health records vs. medicare claims data: a study of 101,980 linked patients[J]. *BMC Neurol*, 2023, 23(1): 325. DOI: [10.1186/s12883-023-03361-w](https://doi.org/10.1186/s12883-023-03361-w).
 - 43 Choi YC, Zhang D, Tyczynski JE. Comparison between health insurance claims and electronic health records (EHRs) for metastatic non-small-cell lung cancer (NSCLC) patient characteristics and treatment patterns: a retrospective cohort study[J]. *Drugs Real World Outcomes*, 2021, 8(4): 577–587. DOI: [10.1007/s40801-021-00269-0](https://doi.org/10.1007/s40801-021-00269-0).
 - 44 Laws MB, Michaud J, Shield R, et al. Comparison of electronic health record-based and claims-based diabetes care quality measures: causes of discrepancies[J]. *Health Serv Res*, 2018, 53 (Suppl 1): 2988–3006. DOI: [10.1111/1475-6773.12819](https://doi.org/10.1111/1475-6773.12819).
 - 45 Carrell DS, Gruber S, Floyd JS, et al. Improving methods of identifying anaphylaxis for medical product safety surveillance using natural language processing and machine learning[J]. *Am J Epidemiol*, 2023, 192(2): 283–295. DOI: [10.1093/aje/kwac182](https://doi.org/10.1093/aje/kwac182).

收稿日期: 2025 年 04 月 15 日 修回日期: 2025 年 05 月 29 日
 本文编辑: 杨 燕 洗静怡