e-ISSN 3050-8460
Journal of Dermatology Research

ATHENAEUM SCIENTIFIC PUBLISHERS

Review Article

Efficiency of Data Extraction from Electronic Medical Records: A Comparative Study in Dermatology

Rogério Nabor Kondo[1*], Liviann de Souza Oliveira[2], Sofia Gonçalves Mota[2], Giulia Akemi Sakashita[2], Bruno Iria Franco[2], Lucas Lopes Cascione[2], Pabllo Batista Moura[2], Maria Heloisa de Souza Bonfim[2], Gustavo Sato Nakamá[2], Gustavo Kendy Kizima[2], Letícia Tainá Morais Machado[2], Tainá Raquel Fernanda da Silva[2], Marcos Eduardo Zanlorenzi[2], Anna Luiza Yorinori[2], Larissa Moreira de Pontes[2], Bruna Tuma[3]

[1]PhD, Dermatologist, Adjunct Professor of Dermatology at the University Hospital of the State University of Londrina, Paraná, Brazil
[2]Medical student at the State University of Londrina, Londrina, Brazil
[3]MSc, Dermatologist, Assistant Professor of Dermatology at the University Hospital of the State University of Londrina, Paraná, Brazil
***Correspondence author:** Rogério Nabor Kondo, PhD, Dermatologist, Adjunct Professor of Dermatology at the University Hospital of the State University of Londrina, Paraná, Brazil; Email: kondo.dermato@uel.br

Citation: Kondo RN, et al. Efficiency of Data Extraction from Electronic Medical Records: A Comparative Study in Dermatology. J Dermatol Res. 2026;7(2):1-5.

https://doi.org/10.46889/JDR.2026.7206

Received Date: 31-05-2026

Accepted Date: 22-06-2026

Published Date: 29-06-2026

Abstract

Background: In 2020, a university outpatient clinic implemented electronic health records, expanding the system to dermatology department in 2022 to centralize data and support research. This study analyzed the tool's impact on scientific support and mapped the epidemiological profile of the dermatology service.

Methodology: This retrospective, observational study (January 2022 to December 2024) compared two methods for extracting clinical and demographic data: manual analysis (conducted by 27 medical students) and automated software extraction (performed by a single IT technician).

Results: A total of 21,432 consultations were conducted across 2,998 patients. The mean age was 50.3 ± 22.0 years (range: 0-97 years), with a female predominance (37.8% male vs. 62.2% female, $p < 0.001$) and most white individuals (64.9%). Psoriasis, non-melanoma skin cancer, atopic dermatitis, actinic keratosis, acne, contact dermatitis, drug eruptions and pruritus were among the most prevalent diagnoses. There were 2,549 surgical procedures, with excision and suture of small lesions or excision and suture of skin and adnexa wounds being the most predominant (62.2%). Data from all 21,432 consultations were provided by the single technician within a single working day, whereas the 27 students reviewed 436 records over 232 days. However, manual analysis captured more refined details, such as exact lesion locations and specific management plans.

Conclusions: Electronic health records systems streamline rapid population-based epidemiological analyses. Nevertheless, automated extraction may overlook detailed clinical nuances, leaving manual review indispensable depending on the research objectives. The dermatological and surgical profile of this tertiary service aligns with national literature and data from the Brazilian Society of Dermatology.

Keywords: Outpatient Care; Dermatology; Skin Diseases; Electronic Health Records; Research

Introduction

The electronic health record system centralizes patient health history and clinical data. Integration with research tools allows cross-referencing data and generates rapid analyses [1,2]. In 2020, the Specialty Outpatient Clinic of the University Hospital of the State University of Londrina (AEHU-UEL) began implementing the electronic health record system (Medview), starting with orthopedics. This was expected to optimize academic research through data cross-referencing. The dermatology specialty adopted digital records in January 2022. However, no studies have yet been conducted in this department using this tool

exclusively to evaluate its actual impact on research.

The objective of this study was to evaluate the impact of implementing electronic health records on the efficiency of scientific research, comparing manual and automated data extraction, as well as mapping the epidemiological profile of dermatological care provided by the institution.

Methodology

This is a retrospective and observational study. The Medical Archive and Statistics Service (SAME) provided records of patients treated at the dermatology outpatient clinics between January 01, 2022 and December 31, 2024. Data collection was performed by medical students and an IT technician. The students manually extracted specific information from each electronic health record, while the technician used software linked to the Medview system to obtain the same data.

To enable access, the clinical directorate of the University Hospital (HU) granted a Medview password to the students of the Academic League of Dermatology (LAD). The data search was based on the International Classification of Diseases (ICD) for dermatological conditions (such as ICD L70.0, corresponding to acne). Each patient had a unique registration number at the HU, although they could have multiple visits and different ICD codes on different dates (for example: psoriasis - L40; ingrown nail - L60 and basal cell carcinoma - C44).

The collected information was entered into the Google Forms platform and subsequently compiled in Microsoft Excel. Statistical analysis was performed using Jamovi (version 2.3.28) and IBM SPSS Statistics (version 22.0). Quantitative variables were described by the mean, minimum and maximum values and standard deviation. The student's t-test was used to compare the means of normally distributed variables. The Z-test of proportions assessed the probability of an event occurring on a given day. Categorical data were expressed as absolute frequencies and percentages. A significance level of $p < 0.05$ was set and 95% Confidence Intervals (CIs) were calculated. The study was approved by the Research Ethics Committee of the State University of Londrina (UEL) under CAAE No. 5766624.7.0000.5231.

Results

Table 1 shows the main demographic characteristics of the study. A total of 2,998 patients generated 21,432 consultations, with a mean age of 50.3 ± 22.0 years, ranging from 0 to 97 years, predominantly female (37.8% male vs. 62.2% female, $p < 0.001$) and of white ethnicity (64.9%) (Table 1).

Characteristics	Values
Age (in years), Mean (SD)	50.3 (22.0)
Minimum-maximum	0 - 97
Gender, %	
Male	37.8
Female	62.2
Ethnicity, %	
White	64.9
Dermatoses, n (%)	
Psoriasis	373 (12.4%)
Non-melanoma cancer€	305 (10.2%)
Atopic dermatitis	209 (7.0%)
Actinic keratosis	208 (6.9%)
Acne	156 (5.2%)
Contact dermatitis	149 (5.0%)
Drug eruptions	126 (4.2%)
Pruritus§	109 (3.6%)

Hidradenitis suppurativa	97 (3.2%)
Others ‡	1,266 (42.2%)
Surgical procedures, n (%) *	2,549
Excision and suturing of small lesions **	1,585 (62.2%)
Skin/soft tissue biopsy	779 (30.5%)
Others #	185 (7.3%)

n = 2,998 patients who generated 21,432 consultations.
SD Standard deviation.
€ denotes Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC)
§ only symptoms of itching (no defined etiology of the itching)
‡ denotes other dermatoses, each representing less than 3%, but which, when added together, accounted for 42.2% of the total.
* There were patients who underwent two or more surgical procedures.
** includes excisions for skin cancer.
#Includes earlobe reconstruction; cutaneous electrocoagulation; lesion removal by shaving; canthoplasty; surgical treatment of blepharochalasis; skin grafting; flap reconstruction.

Table 1: Clinical and demographic characteristics (n = 2998).

Twenty-seven LAD students acted as data abstractors, reviewing 436 medical records over 232 days. Meanwhile, data for all 21,432 clinical encounters were generated by a single technician within one workday. We applied a Z-test of proportions to evaluate the likelihood of an event occurring on a specific day relative to the 232-day group behavior. The p-value indicates the probability of encountering that specific daily result purely by chance, assuming identical population rates ($p < 0.001$).

There was no case-by-case comparison with the ICD in the 436 medical records from the manual analysis (clinical and surgical), which prevented the calculation of the error rate and discrepancy. However, when comparing random records from the manual group with the software data, an underdiagnosis of secondary ICD was found in the system. The manual record proved to be more complete in both clinical and surgical evaluations.

Diagnoses ranged from "A" to "X" (from Acanthosis nigricans/ICD: L83 to Xerosis/L853). Psoriasis, non-melanoma skin cancer, atopic dermatitis, actinic keratosis, acne, contact dermatitis, drug eruptions, pruritus and hidradenitis suppurativa were among the most prevalent diagnoses (Table 1).

A total of 2,549 surgical procedures were performed. Skin/soft tissue biopsy (30.5%) and excision and suture of small lesions or excision and suture of skin and appendage wounds (the most frequently performed at 62.2%) were the most prevalent. Others included: excision of skin tumor/cyst/lipoma; earlobe reconstruction; excision and suture of nevi; electrocoagulation; shaving lesion removal; canthoplasty; surgical treatment of blepharochalasis; surgery and suture with skin Z-plasty; skin excision and grafting and flap reconstruction (Table 1).

Discussion

Electronic medical record systems store patients' complete health history, centralizing clinical data and complementary exams. Their integration with research tools enables data location, cross-referencing and rapid analysis generation, as observed in the present study [3,4].

However, electronic health records may contain gaps that prevent automated data cross-referencing, such as free-text descriptions, unstructured management details, a lack of standardized clinical terminologies and limitations in International Classification of Diseases (ICD) codes. For instance, the ICD code C44 indicates a malignant neoplasm of the skin but fails to specify whether it is recurrent basal cell carcinoma information that would only appear in the header of the outpatient progress notes. In these scenarios, manual data collection becomes unavoidable [3].

Furthermore, medical students have limited hours to dedicate to research. Balancing lectures, tutorials, independent study time and rest periods restricted the daily time available for data gathering. Structural factors, such as password limitations and the requirement to access the software exclusively via physical computers within the hospital, also constrained the process. Table 2 presents the main differences, advantages and disadvantages between the evaluated data extraction strategies.

Characteristics	Manual Extraction	Software Linked to Medview
N.of medical records	436*	2,998
N. of patient visits	3,299*	21,432
N. of abstractors €	27*	1
Time (days) §	232*	1
Advantages	More complete information such as post-operative results and type of treatment.	Faster information; Requires a single computer; Requires fewer data abstractors.
Disadvantages	Time-consuming process; Laborious (requires compiling data into another spreadsheet); Requires the availability of multiple computers; Requires several people.	Written notes cannot be extracted from the database.
* Manual data collection was discontinued when it reached 14.5% of the total patient sample and the 436 records collected manually were a consecutive selection. € denotes the number of students who collected the information manually. § denotes the number of days elapsed to obtain results.		

Table 2: Differences between data extraction methods.

Due to the projected excessively long timeframe for manual verification, data collection via this method was halted after the seventh month (232 days), having completed only 14.5% of the patient sample. Regarding clinical findings, the criteria for dispensing specialized high-cost medications Laudo de Medicamentos Especializados (LME) for specific pathologies may have influenced the prevalence of certain dermatoses. Examples include the supply of isotretinoin for acne management and immunobiologics for psoriasis and hidradenitis suppurativa [5].

Among the dermatitis group, atopic dermatitis and contact dermatitis were among the most common conditions. Actinic keratosis, as a premalignant lesion and non-melanoma skin cancers were also among the most prevalent, converging with literature data.

Drug eruptions (pharmacodermia), categorized under ICD-10 codes (such as L27.0 and L27.1), also showed high prevalence. Population aging and the consequent rise in polypharmacy and drug interactions are often hypothesized as driving factors [5]. Nevertheless, the average age of the drug eruption cohort (48.4 ± 6.6 years) was slightly lower than the overall average of the evaluated population (50.3 ± 22.0 years). This can be explained by the inclusion of individuals in the total group with other dermatoses related to extremes of age, such as skin cancer and xerosis cutis.

The excision of small lesions with suturing was one of the most frequently performed surgical procedures; this technique enables histopathological examination to evaluate surgical margins in both benign and malignant cases, in accordance with the literature [5].

The traditional model of extracting clinical data using rigid coding systems (such as the ICD) often misses the semantic nuances of clinical narratives. Today, Natural Language Processing (NLP) and Large Language Models (LLMs) are transforming this landscape by understanding the contextual meaning of free text without relying on rigid mappings, thereby shifting this paradigm and substantially reducing errors and underreporting inherent in manual chart reviews [6]. This study has limitations regarding being monocentric, retrospective and observational. However, it is expected that future research will benefit from

technological innovations capable of overcoming the physical barrier of manual extraction.

Conclusion

The digital system implemented by the hospital proves to be an effective tool for cross-referencing information, optimizing analysis and promoting scientific research. However, manual data analysis remains necessary to complement qualitative and detailed information, depending on the scope of the study. Additionally, the epidemiological profile of dermatoses observed in this investigation aligned with the patterns described in current scientific literature.

Conflict of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial or non-profit sectors.

Acknowledgement

The authors have no acknowledgments to declare.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Ethical Statement

The protocol was approved by the Research Ethics Committee of the State University of Londrina (CAAE: #5766624.7.0000.5231).

Informed Consent Statement

Informed consent was obtained from all participants included in the study.

Authors' Contributions

All authors contributed equally to this paper.

References

1. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. Clin Res Cardiol. 2017;106(1):1-9.
2. Gouda P, Ezekowitz J. Harnessing electronic medical records in cardiovascular clinical practice and research. J Cardiovasc Transl Res. 2023;16(3):546-56.
3. Okorie CL, Gatsby E, Schroeck FR, Ould Ismail AA, Lynch KE. Using electronic health records to streamline provider recruitment for implementation science studies. PLoS One. 2022;17(5):e0267915.
4. Ferreira IG, Almeida CS, Bulcão LA, Ferreira DG, Weber MB, Bonamigo RR. Hospital dermatology: analysis of dermatological consultations in a tertiary teaching hospital. An Bras Dermatol. 2023;98(5):620-34.
5. Miot HA, Penna GO, Ramos AMC, Penna MLF, Schmidt SM, et al. Profile of dermatological consultations in Brazil (2018). An Bras Dermatol. 2018;93:916-28.
6. Saban M, Lutski M, Zucker I, Uziel M, Ben-Moshe D, Israel A, et al. Identifying diabetes related-complications in a real-world free-text electronic medical records in Hebrew using natural language processing techniques. J Diabetes Sci Technol. 2025;19(4):999-1007.