



Distinguishing cardiac catheter ablation energy modalities by applying natural language processing to electronic health records

Jamie Margetta¹  & Alicia Sale^{*1} 

¹ Department of Health Economics & Outcomes Research, Medtronic, Mounds View, MN 55112, USA

* Author for correspondence: Tel.: +1 651 214 6231; alicia.j.sale@medtronic.com

Aim: Catheter ablation is used to treat symptomatic atrial fibrillation (AF) and is performed using either cryoballoon (CB) or radiofrequency (RF) ablation. There is limited real world data of CB and RF in the US as healthcare codes are agnostic of energy modality. An alternative method is to analyze patients' electronic health records (EHRs) using Optum's EHR database. **Objective:** To determine the feasibility of using patients' EHRs with natural language processing (NLP) to distinguish CB versus RF ablation procedures. **Data Source:** Optum® de-identified EHR dataset, Optum® Cardiac Ablation NLP Table. **Methods:** This was a retrospective analysis of existing de-identified EHR data. Medical codes were used to create an ablation validation table. Frequency analysis was used to assess ablation procedures and their associated note terms. Two cohorts were created (1) index procedures, (2) multiple procedures. Possible note term combinations included (1) cryoablation (2) radiofrequency (3) ablation, or (4) both. **Results:** Of the 40,810 validated cardiac ablations, 3777 (9%) index ablation procedures had available and matching NLP note terms. Of these, 22% (n = 844) were classified as ablation, 27% (n = 1016) as cryoablation, 49% (n = 1855) as radiofrequency ablation, and 1.6% (n = 62) as both. In the multiple procedures analysis, 5691 (14%) procedures had matching note terms. 24% (n = 1362) were classified as ablation, 27% as cryoablation, 47% as radiofrequency ablation, and 2% as both. **Conclusion:** NLP has potential to evaluate the frequency of cardiac ablation by type, however, for this to be a reliable real-world data source, mandatory data entry by providers and standardized electronic health reporting must occur.

Plain language summary: **What was the aim of this research?:** A proof-of-concept study to determine the feasibility of using patients' electronic health record (EHRs) with natural language processing (NLP) to distinguish cryoablation versus radiofrequency ablation procedures across the US.

How was the research carried out?: A retrospective analysis of existing de-identified EHR data using Optum® de-identified EHR dataset and Optum® Cardiac Ablation NLP Table. Medical codes were used to create an ablation validation table. Frequency analysis was used to assess ablation procedures and their associated note terms (Cryoablation, Radiofrequency).

What were the results?: Of the 40,810 validated cardiac ablations, 3777 (9%) index ablation procedures had available and matching NLP note terms. Of these, 22% (n = 844) were classified as generic ablation, 27% (n = 1016) as cryoablation, 49% (n = 1855) as radiofrequency ablation, and 1.6% (n = 62) as both cryoablation and radiofrequency ablation.

What do the results mean?: NLP has potential to evaluate the frequency of cardiac ablation by type, however, for this to be a reliable real-world data source, mandatory data entry by providers and standardized electronic health reporting must occur.

First draft submitted: 12 April 2023; Accepted for publication: 3 January 2024; Published online: 23 January 2024

Keywords: catheter ablation • cryoballoon • electronic health records • natural language processing • pulmonary vein isolation

Background

Atrial fibrillation (AF) is the most common cardiac arrhythmia worldwide. Age is an important risk factor for AF and the growing incidence and prevalence of AF is correlated to an ageing population [1]. An effective treatment option to restore and maintain sinus rhythm for patients with symptomatic AF is pulmonary vein isolation (PVI) via cardiac catheter ablation [1]. Catheter ablation can be performed using either a cryoballoon (CB) catheter or a radiofrequency (RF) catheter [1]. There is limited real world comparative data of CB and RF in the US population as the current ICD-10-CM procedure and procedural terminology (CPT) codes used to identify cardiac ablation procedures are not able to differentiate between these two energy modalities making the distinction impossible to identify in healthcare billing codes [2]. Furthermore, it is crucial to validate a method for distinguishing between these two catheter ablation techniques because research has demonstrated varying results, such as an elevated risk of cardiac perforation and pulmonary vein stenosis associated with radiofrequency ablation [3].

Due to recent advances, electronic health records (EHRs) are becoming increasingly prevalent for research and quality initiatives [4]. An EHR is an individual patient record stored within a hospital's electronic system [5]. EHRs may include patients' diagnoses, treatment plans, medication records and medical history etc. [5]. Data of this type enables treatments, diagnoses, and outcomes to be studied across large patient cohorts [4]. An alternative method to distinguish between CB and RF ablation procedures is to analyze patient's EHRs using natural language processing (NLP) [6].

NLP is a commonly used 'large dataset' evaluation method across healthcare [6]. NLP algorithms can be utilized to parse unstructured physician notes into structured parts. NLP utilizes linguistics technology to read, interpret, and organize important health data that lives within free-text fields. NLP converts the complex clinical narratives (unstructured) into actionable data and insights (structured) [7]. NLP algorithms can be used for information extraction, to capture meaning, or, undergo syntactic processing, to name only a few [8]. The use of NLP can be challenging given the complexity of medical language, although, despite the drawbacks, it is still used for a broad spectrum of healthcare related purposes [9,10].

Due to the unique capabilities NLP offers, healthcare provider notes within a patient's EHR can be searched and "cryoballoon" and "radiofrequency" ablation terms extracted, enabling ablation procedures to be analyzed on a more granular level than available in medical codes or claims based evidence. The aim of this proof-of-concept analysis was to determine the feasibility of using the NLP table to distinguish distinct CB versus RF cohorts and identify any gaps or lessons learned in NLP analysis that could be applied to future comparative analysis. To meet this goal, this analysis had two objectives. The first was to determine ablation energy modalities of patients' index ablation procedure. The second was to broaden the analysis to all ablation procedures and sequence energy modalities by procedure. It was the aim of this analysis to understand how effective NLP is at distinguishing ablation modalities when claims based medical codes are not available to determine feasibility of future use.

Methods

Data source

This was a retrospective analysis of existing de-identified EHR data and an IRB committee waived the requirement for informed consent. For this analysis two datasets were provided by Optum. The first dataset is an Optum EHR dataset. This EHR dataset is payor agnostic but relies on the facility to be an Optum-EHR platform contributing clinic. Patient's data availability is limited to only visits that occurred at Optum contributing sites though historical documentation is sometimes available. The regional breakdown of the contributing hospitals are as follows, 46% are Midwest region 12% Northeast region, 30% Southern region, 8% Western region, and 3% with region unknown. The EHR dataset is 43% female, 57% male, and 0.05% unknown sex. The second dataset is an Optum generated NLP table. This table was generated with the aim of categorizing cardiac ablation patients and therefore the NLP table inclusion criteria was set to only patients with cardiac ablations. Optum generated the NLP table at the request of Medtronic. The NLP table is also payor agnostic and relies on the patient being in the Optum EHR dataset. Medtronic has an on-going contract with Optum to utilize this database. Standard SQL coding using Cloudera-ODBC Driver for Impala was used for this analysis. Data were also transferred to a SAS 9.4 server for further analytical support.

Table 1. Table linking variable definitions.

Step	Table	Definition	Variables
1	Internal validation	Medtronic Internal Optum EHR table. Data years include 2016–2019	1. Encrypted patient identifier 2. ICD-10-CM diagnostic code 3. ICD-10-PCS or CPT code 4. Procedure date
1	Optum NLP ablations	Optum generated cardiac ablation table. Data years 2016–2019.	1. Patient identifier 2. Note term(s) 3. Note date
2	Link on patient identifier and note date $\pm$ 3 days after procedure date		
3	Validated analytical sample	EHR medical code defined procedures and their respective note terms from Optum NLP table	1. Patient identifier 2. ICD-10-CM diagnostic code 3. ICD-10-PCS or CPT code 4. Procedure date 5. Note term(s) 6. Note date

EHR: Electronic health record; CPT: Current procedural terminology; NLP: Natural language processing.

Development & validation of NLP table

NLP's purpose is to extract key information from EHR records. Optum's NLP system was developed using vocabulary from the Unified Medical Language System (UMLS) that includes multiple medical dictionaries. These include the Logical Observation Identifiers Names and Codes (LOINC), the Systemized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) and RxNorm terms within these dictionaries are distinguished as terms of interest within the NLP System (tokens) so that the presence of these terms in the free text will result in their inclusion in the semi-structured data. Development and validation of NLP programming involves manual review of a sample of narrative texts to identify appropriate targets for the NLP process, and to confirm the correct output. The algorithms require complex computer programming to search text fields and recognize syntax and formatting in EHRs to convert facts into a structured format. The NLP System uses an architecture based on pipelines and standoff annotations as described in the Unstructured Information Management Architecture (UIMA) The specific stages of the pipeline are proprietary to Optum but are like other UIMA based systems described by Savova *et al.* [11]. The primary objective of this analysis was to specifically evaluate pulmonary vein isolation (PVI) for patients with atrial fibrillation (AF). Additionally, patients who presented with atrial flutter (AFL) unrelated to pulmonary vein triggers during their ablation procedure and required a [supplementary](#) SVT ablation were categorized as a distinct group [12,13]. These individuals are referred to as 'SVT add-on ablation' patients. The hypothesis was that, if initially treated with cryoablation, these patients would subsequently need a second radiofrequency catheter ablation to address their AFL, potentially constituting a substantial portion of the broader 'ablation' cohort.

The NLP process does not utilize medical codes for its patient inclusion criteria therefore it was not possible to create these PVI/SVT distinctions or sub-categories of procedure types with the NLP table alone. To both validate the NLP table and create PVI/SVT patient cohorts, an internal validation table was created using CPT codes, ICD-10-CM diagnostic codes, and ICD-10-PCS codes ([Supplementary Tables 1–3](#)) from the overarching EHR data. Coding definitions of PVI for AF ablation and SVT add on ablation can be found in [Supplementary Tables 1–3](#). This internal validation table was then linked to the Optum NLP table on available data variables ([Supplementary Table 4](#)). Patients from the NLP table were linked to the internal validations table by an encrypted patient identifier and their NLP note date must have occurred within three days post their ablation procedure date. Specific healthcare facilities are not identifiable in this dataset therefore it was not possible to trace the procedures back to a single patient ([Table 1](#)).

Index procedures

To analyze patients' index (their first) cardiac ablation procedures, the internal validation table was limited to index ablation procedures only and any accompanying SVT add-on procedures. Due to the structure of Optum EHR data, it was possible a patient's index ablation procedure was not available in the data because it was performed at a non-Optum contributing hospital prior to their inclusion in the Optum dataset. Unlike claims data, EHR data do not have continuous enrollment information to capture a patient's enrollment status. To better assess if the index procedure identified in this study was truly Index, we analyzed patients with a history of an NLP note term greater than 30 days prior to their indicated index procedure.

Table 2. Term combinations.	
1-term combinations	Classification
Cryoablation	Cryoablation
Radiofrequency	Radiofrequency
Ablation	Ablation
2-term combinations	Classification
Radiofrequency + ablation	Radiofrequency
Cryoablation + ablation	Cryoablation
Radiofrequency + cryoablation	Both
3-term combinations	Classification
Ablation + cryoablation + radiofrequency	Both

If a patient had an NLP note term greater than 30 days prior to their indicated index procedure, it is likely the patient had their first ablation procedure at a non-Optum clinic, entered the Optum database on a later date and their NLP terms were referencing historical documentation. For this reason, patients with note terms >30 days prior to their index procedure were excluded for not meeting index criteria. A sensitivity analysis was performed assessing patients with note terms >60 days prior to their index procedure and no meaningful differences were found (Supplementary Data). It is possible for a single procedure to have more than one note term associated. All note terms per procedure were included in this analysis.

Of the remaining patients, a frequency analysis was conducted to determine the NLP note terms associated with their index ablation procedure. The three-note term possibilities were (1) ‘ablation’, (2) ‘cryoablation’, (3) ‘radiofrequency ablation’. Since terms were distinct, repetitions of the same terms were not possible – creating a maximum amount of three terms per procedure. Note, term combinations were then further categorized via the method below (Table 2). Term combinations from all index procedures were summed to categorize the frequency of note terms.

Multiple ablation procedures

To analyze all ablation procedures, the internal validation table was not limited to index procedures. Therefore, patients could contribute multiple ablation procedures. Note, SVT ablation for atrial flutter is defined as an add-on procedure and must be accompanied by a PVI for AF procedure indicated on the same day. Sequential procedures required at minimum a PVI for AF code but could include the SVT atrial flutter add on codes if indicated.

It was possible for more than one note term to be associated with a procedure. All note terms associated with their respective procedure were included. It was also possible a patient could have sequential ablation procedures within three days of each other. Frequency analysis was performed on procedures that occurred within three days of each other to determine if the note terms associated with both procedures were the same. Out of the procedures that occurred within three days of each other, 100% of terms were categorized as the same terms, and therefore only the initial procedure and note terms were retained. Any procedures that occurred outside of a three-day window were included as sequential procedures. The maximum number of ablation procedures a single patient had was four procedures. The same process used to categorize the index procedures was used to categorize frequency of note terms per procedure in the multiple analysis (Table 2).

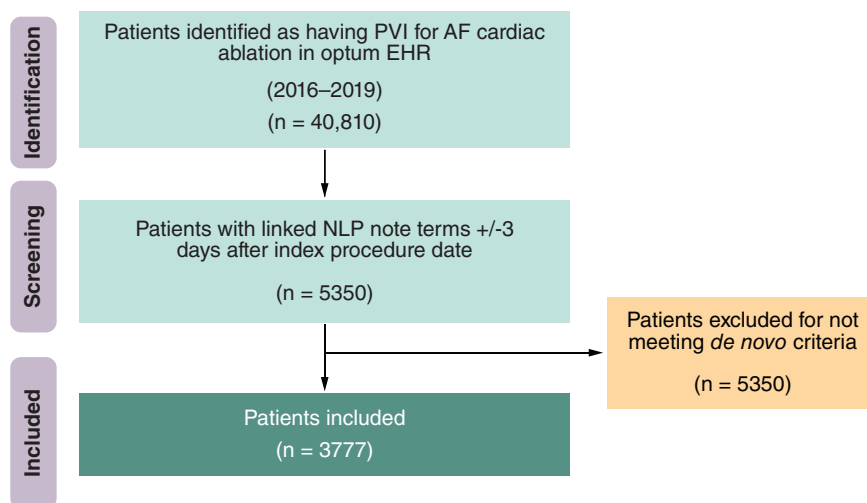
Results

Index procedures

Data availability was limited to the number of index ablation procedures that could be identified and linked to a note term(s). b. Of the 40,810 patients identified with medical codes only 13% (n = 5350) had linked note terms within three days after their procedure date (Table 1). The patient baseline characteristics are displayed in Table 3. Furthermore, another 1573 patients were excluded for likely having an index ablation procedure at a non-Optum contributing hospital (Figure 1). The remaining analytical sample size for the index cohort was then 9% (n = 3777) (Figure 1). This 9% represents the proportion of cardiac ablation procedures identified via medical codes who had available note terms in the NLP table. This indicates many procedures do not utilize note documentation.

Table 3. Characteristics of patients with at least one procedure (n = 5350).

Sex	n	%
Female	1803	33.7%
Male	3544	66.2%
Unknown	3	0.1%
Race		
Caucasian	4838	90.4%
Asian	90	1.7%
African–American	211	3.9%
Other/unkown	211	3.9%
Region		
West	598	11.2%
South	805	15.0%
Northeast	2289	42.8%
Midwest	1539	28.8%
Other/unknown	119	2.2%
Age ranges (years)		
<=30	14	0.3%
31–40	52	1.0%
41–50	177	3.3%
51–60	647	12.1%
61–70	1688	31.6%
71–80	1865	34.9%
81–90	828	15.5%
>91	79	1.5%

**Figure 1. Index procedures screening and inclusion.**

AF: Atrial fibrillation; EHR: Electronic health record; NLP: Natural language processing; PVI: Pulmonary vein isolation.

Frequency of note terms

Of the 3,777 index ablation procedures with available NLP note terms, 22% (n = 844) had an ‘ablation’ term associated, 27% (n = 1016) had a ‘cryoablation’ term associated, 49% (n = 1855) had a ‘radiofrequency’ term associated, and 2% (n = 62) had both ‘cryoablation’ and ‘radiofrequency’ terms associated (Figure 2).

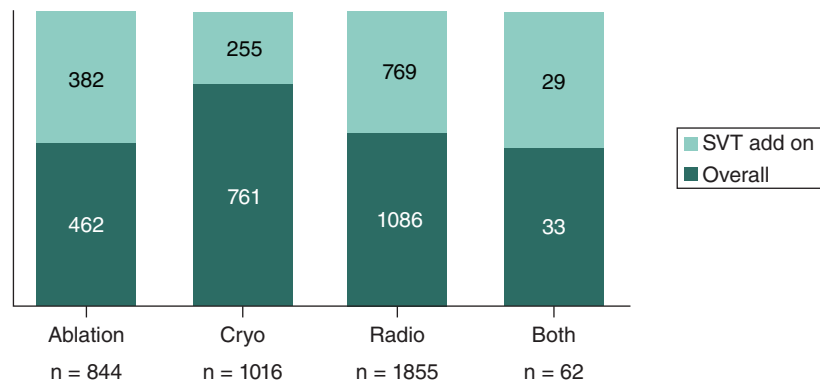


Figure 2. Frequency of note terms and supraventricular tachycardia add on codes across index procedures.
SVT: Supraventricular tachycardia.

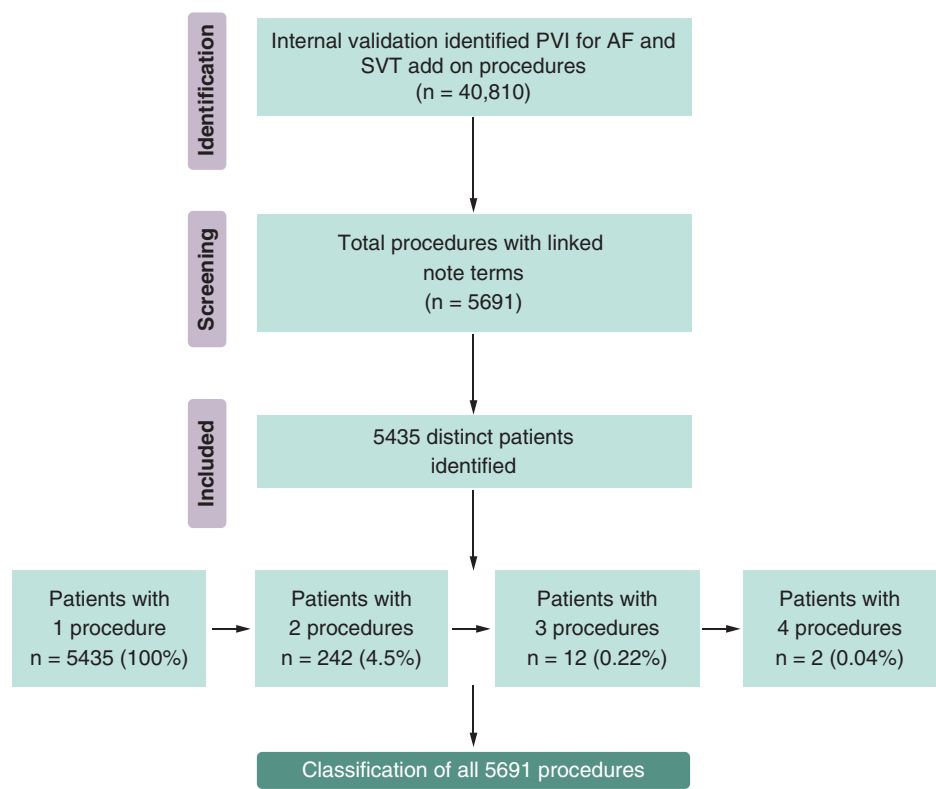


Figure 3. Multiple procedures screening and inclusion.
SVT: Supraventricular tachycardia.

Frequency of SVT-add on procedures

Of the 844 ablation terms, 45.3% (n = 382) were categorized as SVT add-on procedures. Of the 1,016 cryoablation terms, 25.5% (n = 255) were categorized as SVT add-on procedures, and of the 1855 radiofrequency terms, 41.5% (n = 769) were categorized as SVT add-on procedures (Figure 2).

Multiple procedures

For the multiple procedures analysis, patients were allowed to contribute more than one ablation procedure. Of the 40,810 cardiac ablations identified with medical codes, 14% (n = 5691) had note terms available within three days after a procedure date (Figures 3 & 4) in the NLP table. The slight increase in patient count with available

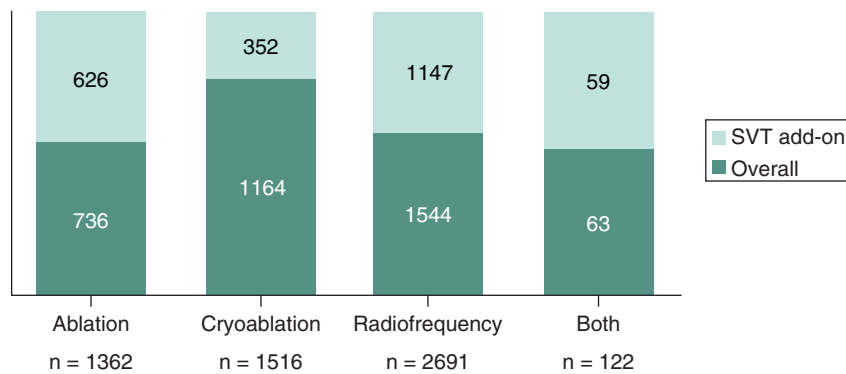


Figure 4. Frequency of note terms and supraventricular tachycardia add on codes across multiple procedures.
SVT: Supraventricular tachycardia.

note terms is likely due to a patient's index ablation not meeting inclusion criteria, but a later procedure date did meet inclusion criteria and had associated terms found in the NLP table.

Frequency of note terms

The final multiple procedures with available notes in the NLP table had 5435 patients and 5691 ablation procedures. 4.5% (n = 242) of all patients had two ablation procedures, 0.22% (n = 12) had three ablation procedures and 0.04% (n = 2) had four ablation procedures (Figure 3). Of the 5691 ablation procedures with available NLP not terms, 24% (n = 1362) had an 'ablation' term associated, 27% (n = 1516) had a 'cryoablation' term associated, 47% (n = 2691) had a 'radiofrequency' term associated, and 2% (n = 122) had both 'cryoablation' and 'radiofrequency' terms associated (Figure 4).

Frequency of SVT-add on procedures

Of the 1362 ablation terms, 46% (n = 626) were categorized as SVT add-on procedures. Of the 1516 cryoablation terms, 23.2% (n = 352) were categorized as SVT add-on procedures, of the 2691 radiofrequency terms, 43% (n = 1147) were categorized as SVT add-on procedures, and of the 122 both terms, 4.8% (n = 59) were categorized as SVT add-on procedures (Figure 4).

Discussion

In this feasibility study, we established that employing NLP on de-identified EHR effectively distinguishes between CB and radiofrequency ablation methods. Our findings indicated that it is possible to distinguish ablation modality types using NLP where billing codes are absent.

Utilizing the notes from the NLP table, our results showed that radiofrequency ablation was the prevailing choice for both initial and repeated ablation procedures, accounting for 49% and 47%, respectively. This outcome is not unexpected, given the longer history of radiofrequency ablation and its focal design that allows it to target cardiac areas beyond the pulmonary veins. Due to the novelty of using NLP, it proves challenging to directly compare our findings with other studies that examine the prevalence of CB and radiofrequency ablation. However, these findings do align with those from the 2020 Get With the Guidelines Ablation real-word registry in the USA, which also revealed that irrigated radiofrequency ablation was the predominant energy modality (70.5%) nationwide, followed by CB ablation (23.7%) [14].

The unique method of pairing the medical codes in EHR with the NLP table also highlights the ability to distinguish procedures with SVT add-on codes. Our results illustrated that for the terms 'ablation' and 'both' cohorts, nearly half of the ablation procedures had an SVT add-on code. This could illustrate that these patients exhibited non-pulmonary vein triggers caused by atrial flutter (AFI) during their AF ablation procedure requiring an additional SVT ablation [12–14]. This may explain why a physician would use 'both cryoballoon and radiofrequency' in one ablation procedure. Moreover, it may also explain why a non-specific 'ablation' term was captured in the EHR because both catheters were used. However, the other half of the procedures were not SVT and thus portrays the uncertainty of non-descriptive note terms written by providers in patients EHRs.

There are multiple sources of real-world data (RWD) including, claims, registries and EHR [15,16]. Of all these, EHR has the greatest potential to understand potential differences in health outcomes as it includes a plethora of clinical information such as narrative text notes, procedures, medications, diagnoses etc. [15,16]. The clinical notes are particularly valuable because they capture many complexities of patient encounters and outcomes that are underreported or absent in billing codes. This study showed utilizing NLP on HER clinical notes was successful in identifying discrete cohort for CB and RF for retrospective analysis when conducting a comprehensive administrative claims analysis was not feasible due to the absence of distinct coding. The NLP tool efficiently processed a substantial EHR dataset; nevertheless, utility was restricted by availability of note terms in the NLP ablation table and healthcare providers documenting notes within the patients' EHRs.

Drawing definitive conclusions from this outcome is challenging due to limitations imposed by unequal data quality and the lack of notes that differentiate ablation energy types. The active involvement of healthcare providers in documenting note terms within patients' EHRs is essential for distinguishing between CB and RF when employing NLP techniques. This highlights the need for standardized electronic health reporting across providers. Our study revealed that less than 15% of cardiac ablation patients had notes accessible in the NLP table. This discrepancy could be attributed to multiple challenges associated with achieving standardized electronic reporting such as, variations in provider documentation practices and the diverse content and field logic/rules within EHR data entry forms, often resulting in incomplete data that varies according to the healthcare practice and EHR vendor [17]. Even when multiple healthcare systems utilize the same vendor, they may have different data entry rules and validation procedures [17–19]. Moreover, the significant documentation burden placed on physicians, who already have limited time, coupled with the absence of provider incentives to specify the energy modality in a patient's EHR, primarily because it's not mandatory, further compounds the challenge [17–20]. Potential solutions to address this lack of standardization is firstly, to implement common data standards and terminology across the USA, second, establish regulatory mandates and incentives for standardized EHR reporting and lastly, encourage EHR vendors to work together to create interoperable EHR systems and standard reporting templates.

Limitations

Our study has several limitations. First, it is limited to only hospitals with an Optum contributing platform. A large portion of these hospitals (46%) are in the Midwest and a very small portion (3%) in the Western region. Potential differences in regional locations should be considered as a limitation of this analysis. Moreover, only 9% (n = 3777) of cardiac ablation procedures identified via medical codes had available note terms in the NLP table. This limited our small sample size and indicated many procedures do not utilize documentation. It is possible there are meaningful differences in patients with and without documentation. No additional analysis was undertaken to consider these potential differences and future work should seek to understand these differences. The small sample size may also limit the results from being extrapolated beyond this analysis.

Second, NLP methods do not utilize medical codes and therefore as Optum created the NLP table a second-stage internal validation was required. As previously stated, the internal validation table relied on ICD-10-CM diagnostic and procedural codes. This validation was time consuming and limited the use of the NLP table to only those patients that had a matched link thus, limiting generalizability. Third, there were limitations surrounding medical codes used. No inpatient PVI specific ablation code existed, limiting the specificity. Moreover, EHR and NLP data are inherently limited by their completeness and quality [5,8,9]. With varying provider practices, if specific words, terms, or notes are not regularly input into the patient's record, these data will be absent from the NLP table as seen in these analyses [5,8,9].

Conclusion

In summary, our analyses highlight the potential for NLP to evaluate the frequency of different types of cardiac catheter ablation, however, for this to be a reliable RWD in the future, mandatory data entry by providers must occur. This is of particular importance given the additional ablation energy modalities due to launch in the USA over the next few years. Understanding differences in health outcomes and quality measures across energy modalities will be crucial in improving healthcare.

Summary points

- There is limited real world data of cryoballoon (CB) and radiofrequency (RF) cardiac ablation procedures in the US as healthcare codes are agnostic of energy modality.
- An alternative method is to analyze patients' electronic health records (EHRs) using Optum's EHR database.
- To determine the feasibility of using patients' EHRs with natural language processing (NLP) to distinguish CB versus RF ablation procedures.
- This was a retrospective analysis of existing Optum® de-identified EHR dataset, Optum® Cardiac Ablation NLP Table. Medical codes were used to create an ablation validation table.
- Frequency analysis was used to assess ablation procedures and their associated note terms.
- Of the 40,810 validated cardiac ablations, 3777 (9%) index ablation procedures had available and matching NLP note terms. Of these, 22% (n = 844) were classified as generic 'ablation', 27% (n = 1016) as cryoablation, 49% (n = 1855) as radiofrequency ablation, and 1.6% (n = 62) as both.
- NLP has potential to evaluate the frequency of cardiac ablation by type, however, for this to be a reliable real-world data source, mandatory data entry by providers and standardized electronic health reporting must occur.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: <https://bpl-prod.literatumonline.com/doi/10.57264/cer-2023-0053>

Author contributions

Both authors were responsible for study conception, design and drafting the manuscript. Jamie Margetta was responsible for analyzing the data.

Financial disclosure

The authors have no financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Competing interests disclosure

J Margetta and A Sale are both employees and stockholders of Medtronic. The authors have no other competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript apart from those disclosed.

Writing disclosure

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state they have obtained institutional review board exemption from WCG IRB's IRB Affairs Department for the research described because identity of subjects cannot be identified.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

References

Papers of special note have been highlighted as: • of interest

1. Hindricks G, Potpara T, Dagres Net *et al.*, ESC Scientific Document Group *et al.* 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur. Heart J.* 42(5), 373–498 (2021).
 2. Coding and Payment Guide. Cardiac catheter ablation. Economics, reimbursement, and evidence (2021). <https://asiapac.medtronic.com/content/dam/medtronic-com/us-en/hcp/reimbursement/documents/catheter-ablation-coding-reimbursement-guide.pdf>
- **Cardiac ablation codes were sourced from here.**

3. Hachem AH, Marine JE, Tahboub HA *et al.* Radiofrequency ablation versus cryoablation in the treatment of paroxysmal atrial fibrillation: a meta-analysis. *Cardiol. Res. Pract.* 2018, 6276241 (2018).
4. Shah RU, Mukherjee R, Zhang Y *et al.* Impact of different electronic cohort definitions to identify patients with atrial fibrillation from the electronic medical record. *J. Am. Heart Assoc.* 9(5), e014527 (2020).
- **Similar analysis in arial fibrillation therapy area.**
5. Laaksonen N, Varjonen JM, Blomster M *et al.* Assessing an electronic health record research platform for identification of clinical trial participants. *Contemp. Clin. Trials Commun.* 21, 100692 (2020).
6. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: a systematic review. *Int. J. Med. Inform.* 114, 57–65 (2018).
7. Optum. Using AI to solve health care's real problems. Natural Language Processing (2021).
<https://www.optum.com/business/resources/ai-in-healthcare/nlp-ai-roi.html>
8. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J. Am. Med. Inform. Assoc.* 26(4), 364–379 (2019).
9. Fleuren WWM, Alkema W. Application of text mining in the biomedical domain. *Methods* 74, 97–106 (2015).
10. Wang Y, Wang L, Rastegar-Mojarad M *et al.* Clinical information extraction applications: a literature review. *J. Biomed. Inform.* 77, 34–49 (2018).
11. Savova GK, Masanz JJ, Ogren PV *et al.* Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J. Am. Med. Inform. Assoc.* 17(5), 507–513 (2010).
12. Pérez FJ, Schubert CM, Parvez B, Pathak V, Ellenbogen KA, Wood MA. Long-term outcomes after catheter ablation of cavo-tricuspid isthmus dependent atrial flutter: a meta-analysis. *Circ. Arrhythm Electrophysiol.* 2(4), 393–401 (2009).
13. Della Rocca DG, Di Biase L, Mohanty S *et al.* Targeting non-pulmonary vein triggers in persistent atrial fibrillation: results from a prospective, multicentre, observational registry. *Europace* 23(12), 1939–1949 (2021).
14. Loring Z, Holmes DN, Matsouaka RA *et al.* Procedural patterns and safety of atrial fibrillation ablation: findings from get with the guidelines-atrial fibrillation. *Circ. Arrhythm Electrophysiol.* 13(9), e007944 (2020).
15. US Food and Drug Administration. Sentinel System – Five-Year Strategy 2019–2023 (2019).
<https://www.fda.gov/media/120333/download>
16. Dhruva SS, Ross JS, Desai NR. Real-world evidence: promise and peril for medical product evaluation. *PT* 43(8), 464–472 (2018).
- **Outlines real-world evidence possibilities.**
17. Divney AA, Lopez PM, Huang TT, Thorpe LE, Trinh-Shevrin C, Islam NS. Research-grade data in the real world: challenges and opportunities in data quality from a pragmatic trial in community-based practices. *J. Am. Med. Inform. Assoc.* 26(8–9), 847–854 (2019).
18. Savitz ST, Savitz LA, Fleming NS, Shah ND, Go AS. How much can we trust electronic health record data? *Healthc. (Amst.)* 8(3), 100444 (2020).
19. Johnson SB, Bakken S, Dine D *et al.* An electronic health record based on structured narrative. *J. Am. Med. Inform. Assoc.* 15(1), 54–64 (2008).
20. Callahan A, Fries JA, Ré C *et al.* Medical device surveillance with electronic health records. *NPJ Digit. Med.* 2, 94 (2019).