

ChartSweep: A HIPAA-compliant Tool to Automate Chart Review for Plastic Surgery Research

Christian Chartier
Lisa Gfrerer, MD, PhD
William G. Austen, Jr, MD

Summary: Retrospective chart review (RCR) is the process of manual patient data review to answer research questions. Large and heterogeneous datasets make the RCR process time-consuming, with potential to introduce errors. The authors therefore designed and developed ChartSweep to expedite the RCR process while remaining faithful to its methodological rigor. ChartSweep is an open-source tool that can be customized for use with any electronic health record system. ChartSweep was developed by the authors to extract information from electronic health records using the Python coding language. As proof-of-concept, the tool was tested in three studies: RCR1—Identification of subjects who underwent radiofrequency ablation in a cohort of patients who had undergone headache surgery (n = 172); RCR2—Identification of patients with a diagnosis of thoracic outlet syndrome in patients who underwent peripheral neuroplasty (n = 806); RCR3—Identification of patients with a history of implant illness or breast implant-associated anaplastic large cell lymphoma in patients who had undergone implant-based breast augmentation or reconstruction (n = 1133). Inter-rater reliability was assessed. ChartSweep reduced the time required to conduct RCR1 by 1315 minutes (21.9 hours), RCR2 by 1664 minutes (27.7 hours), and RCR3 by 2215 minutes (36.9 hours). Inter-rater reliability was uncompromised (k = 1.00). Open-source Python libraries as leveraged by ChartSweep significantly accelerate the RCR process in plastic surgery research. Quality of data review is not compromised. Further analyses with larger, heterogeneous study populations are required to further validate ChartSweep as a research tool. (*Plast Reconstr Surg Glob Open* 2021;9:e3633; doi: 10.1097/GOX.0000000000003633; Published online 15 June 2021.)

INTRODUCTION

Retrospective chart review (RCR) is the process of manual patient data review to answer research questions. Although widely used in peer-reviewed clinical studies, there is no consensus on the best method of conducting RCR.^{1,2} In its original form, the RCR process involves data extraction using pen and paper from a physical chart. Poor quality control and inter-rater variability/ subjectivity are disadvantages of this form of RCR and are compounded in studies with large patient populations and heterogeneous data.³ The advent of electronic health records (EHRs) and a wide array of advanced data extraction software

packages has shifted modern RCR to the electronic setting. The value proposition of EHR management systems is to efficiently and safely document patient and disease progress, support disease management, facilitate coding for research and billing and ease provider-patient and inter-provider communication.^{4,5} RCR in the EHR setting is more centralized, and cost-effective, and less error-prone. Nonetheless, lack of standardization remains a flaw of RCR in the current technological environment.⁵

Certain research variables such as laboratory values and other numeric test results are easier to interpret with RCR than operator-dependent and heterogeneous variables such as surgical reports/clinic notes/free text. This makes RCR particularly difficult in surgical disciplines, where large databases can be unstructured, with pertinent clinical information buried in plain text narrative. Recently, data scientists versed in natural language processing (NLP), a sub-field of machine learning and artificial intelligence (AI), have proposed applications to more

From the Division of Plastic and Reconstructive Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Mass.

Received for publication January 25, 2021; accepted April 14, 2021.

Copyright © 2021 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/GOX.0000000000003633

Disclosure: All the authors have no financial interest to declare in relation to the content of this article.

Related Digital Media are available in the full-text version of the article on www.PRSGlobalOpen.com.

easily analyze EHRs.^{6–10} These platforms leverage data from millions of patient files to interpret medical language and reach meaningful conclusions with less time spent reviewing individual EHRs. However, these innovative uses of technology have largely not yet reached commercialization.¹¹ Therefore, there is a big need to rethink RCR methodology as we use it today to process large heterogeneous datasets and produce reliable outputs/insights fast. Streamlining RCR in surgical disciplines will allow more time to be spent on study design and data analysis.

The authors therefore designed and developed ChartSweep, a HIPAA-compliant Windows (Microsoft Corporation, Wash.) and Mac (Apple Inc., Calif.) application leveraging the Python coding language to streamline and expedite the RCR process while remaining faithful to its methodological rigor as outlined by Matt and Matthew.^{5,12} ChartSweep is a free tool available to researchers upon request and can be customized for use with any EHR system, though it has currently only been used on Epic EMR (Epic Systems Corporation, Wis.).

METHODS

We performed three RCR studies with increasing patient numbers: RCR 1—identification of subjects who underwent radiofrequency ablation in a cohort of patients who had undergone trigger site deactivation surgery (n = 172); RCR 2—identification of patients with a diagnosis of TOS in patients who underwent peripheral neuroplasty (n = 806); RCR 3—identification of patients with a history of implant illness or breast implant-associated anaplastic large cell lymphoma (BIA-ALCL) in patients who had undergone implant-based breast augmentation or reconstruction (n = 1133).

All three retrospective chart reviews were approved by the Institutional Review Board at the Massachusetts General Hospital.

ChartSweep Development

ChartSweep is a tool developed at the Division of Plastic and Reconstructive Surgery, Massachusetts General Hospital. ChartSweep was coded in the Python programming language. It uses the Selenium (<https://www.selenium.dev/>) and Pynput (<https://pypi.org/project/pynput/>) Python libraries to extract information from EHRs and securely store it in .csv, .txt, .pdf or .jpeg format. These libraries—freely-accessible fragments of pre-written code—allow developers to automate computer tasks by using code to manipulate mouse and keyboard functions. (See **table 1, Supplemental Digital Content 1**, which displays the Selenium Python library sample code. (<https://www.selenium.dev/documentation/en/introduction/>). <http://links.lww.com/PRSGO/B674>.) (See **table 2, Supplemental Digital Content 2**, which displays the Pynput Python library sample code (<https://pypi.org/project/pynput/>). This sequence allows the user to manipulate a computer's mouse to automate a task. <http://links.lww.com/PRSGO/B675>.)

ChartSweep has the ability to search through all components of the EHR (clinical/surgical notes, laboratory results, imaging study reports, etc.) to identify a term/diagnosis/

complication/laboratory result of interest. If a patient record contains the queried value, ChartSweep records the MRN/context and appends them to an output list (.txt) for manual review. Further, ChartSweep can generate a list of MRNs of patients who underwent a surgical procedure using a list of current procedural terminology (CPT) codes.

ChartSweep's HIPAA-compliance relies on the principles of access control, audit control, and information control:

1. Access control: A user deploying ChartSweep to extract information from the EHR must “log into” the EHR using their unique username and password as they would during manual review. ChartSweep can only be deployed on encrypted workstations with EHR access.
2. Audit control: All attempts at accessing protected health information are logged by the EHR, regardless of ChartSweep use. Importantly, the user must provide ChartSweep with a list of medical record numbers before beginning the search. As with manual review, all patients on this list must be part of an institutional review board–approved study. Importantly, ChartSweep must be reinitialized every 15 minutes to prevent automatic log off after prolonged periods of inactivity. This ensures the EHR user must remain at the workstation throughout the data extraction process.
3. Information storage: ChartSweep is configured to extract and store information on encrypted platforms in compliance with data safety protocols outlined by our institutional review board.

Retrospective Chart Reviews

As a first proof-of-concept, a RCR of 172 patient records stored in Epic EMR (Epic Systems Corporation, Wis.) was performed to identify subjects who had undergone radiofrequency ablation (RFA) of the greater or lesser occipital nerves (GONs/LONs) before trigger site deactivation surgery for treatment of headaches. First, a clinical researcher conducted the RCR manually according to standard methodology.⁵ Then, a second automated RCR was conducted utilizing ChartSweep. In this context, ChartSweep scanned for the following terms: “ablation,” “radiofrequency,” “radio” and “RFA.” Automated ChartSweep output was then reviewed and patient charts describing RFA in other contexts (lumbar ablation, endometrial ablation) were manually excluded. Total time required for each review (timed manual review versus ChartSweep time to comparable output) was recorded, and discrepancies between data output were evaluated using inter-rater reliability (ChartSweep versus manual RCR).

ChartSweep was then deployed to identify patients with a confirmed diagnosis of thoracic outlet syndrome (TOS) from a cohort of patients who underwent upper extremity neuroplasty between 8/2011 and 3/2020. A dataset of 806 patient records was generated from the Partners' Health Care Research Patient Data Repository using the CPT billing code for peripheral neuroplasty (64708). ChartSweep

used the specific terms “TOS,” “outlet,” and “thoracic” as well as the non-specific term “syndrome” to identify diagnoses of TOS. A sample of 20 charts was reviewed by a trained clinical researcher to determine time spent for review and inter-rater reliability.

Lastly, ChartSweep was used to define a cohort of patients who underwent implant-based breast reconstruction or augmentation between April 2016 and March 2020 (CPT codes 19340, 19342, 19370) and who presented with symptoms or a documented history of implant illness or BIA-ALCL. The terms “ALCL,” “lymphoma,” “CD30,” “fatigue,” “confusion,” “swelling,” “weight gain,” “weight loss,” and “implant illness” were used, as these terms were found to be associated with both as published in the BIA-ALCL Patient Advisory American Society of Plastic Surgeons position statement and safety advisory.^{13,14} A sample of 20 charts was reviewed by a trained clinical researcher to determine time spent for review and inter-rater reliability.

RESULTS

Radiofrequency Ablation

Total time spent on manual review of 172 patient records was 1371 minutes (22.9 hours), with a mean evaluation time per medical record of 8 minutes. Automated ChartSweep review was significantly faster, requiring 56 minutes overall, and 0.3 minutes per patient record ($P < 0.0001$). Time saved—the difference between manual review time and the time required for ChartSweep to achieve a comparable result—was 7.7 minutes per chart and 1315 minutes (21.9 hours) total (Table 1). Both reviews identified 16 patients who had undergone RFA out of 172 total patients with excellent inter-rater reliability ($k = 1.00$).

Thoracic Outlet Syndrome

ChartSweep reviewed 806 patient charts and correctly identified 432 patients treated for TOS. Automated review time was 109 minutes (1.8 hours), with a mean evaluation time per medical record of 0.1 minutes per patient record. Manual review was performed for 20 patient records with total review time of 43 minutes. Inter-rater reliability was 1.00. Based on manual review time for 20 records, total manual review time was 1773 minutes (28.9 h). Time saved by ChartSweep was 1664 minutes (27.7 hours) (Table 1).

Implant Illness and BIA-ALCL

CPT code review revealed 1133 patients who underwent implant-based breast reconstruction or augmentation

between 4/2016 and 3/2020. The algorithm successfully identified one case of implant illness using the term “implant illness.” Further, 10 mentions of the term “CD30” were identified, all of which were in the context of a previous unrelated diagnosis of lymphoma and were therefore excluded. Seventy-five mentions of “ALCL” were detected, which were manually excluded because the term was used in the contexts of standard surgical consents and to reassure patients at low risk of BIA-ALCL. No cases of BIA-ALCL were identified, consistent with department-wide prospectively maintained logs. Inter-rater reliability (on 20 patient files reviewed manually) was 1.00. Manual review was performed for 20 patient records, with total review time of 42 minutes. Total extrapolated manual review time was 2345 minutes (39.1 hours). Time saved by ChartSweep across 1133 patients was 2215 minutes (36.9 hours).

DISCUSSION

Manual RCR has several limitations, including high inter-rater variability/subjectivity, and long review time in studies with large patient populations and heterogeneous data.³ This study evaluated the utility of ChartSweep, an algorithm developed to expedite the RCR process across small, medium, and large datasets. ChartSweep significantly reduced total RCR time compared with manual RCR ($P < 0.0001$), without compromising methodological rigor. Inter-rater reliability between human review and algorithmic review was excellent ($k = 1.00$ in both proofs-of-concept).

Current database creation and RCR methodology rely heavily on manual review. In large patient cohorts, this practice is time-consuming and can be error prone.^{5,15} Chart Sweep is able to reduce the subjective bias introduced during manual review by objective data compilation. Further, in an era of increasing clinical demands, dedicated research time is sparse.^{16,17} There is a huge need for methods to reduce time spent on manual data review. ChartSweep was able to reduce the time needed to review charts across three RCR studies, resulting in 5194 minutes (86.6 hours, 2.5 week-equivalents for a full-time researcher) saved. By reducing time spent on the repetitive, error-prone components of RCR, the total overall time-to-publication is reduced. Over the course of a research career spanning 35 productive years, the amount of time saved would be significant and researcher productivity could be significantly increased.

As increasingly sophisticated and standardized EHR platforms spanning entire provider/hospital networks are implemented, it is in researchers’ best interests to adopt technologies capable of interpreting these larger datasets.¹⁸ ChartSweep makes studies requiring thorough RCR of large

Table 1. Comparison of ChartSweep and Manual Reviews

Task	No. Patients	Manual Review Time (Min)	ChartSweep Review Time (Min)	Time saved (Min)
Radiofrequency ablation among operative headache patients	172	1371	56	1315
Thoracic outlet syndrome among peripheral neuroplasty patients	806	1773*	109	1664
Implant illness and BIA-ALCL	1133	2345*	130	2215

*Denotes extrapolated total review time based on 20 reabstracted patient records used to determine inter-rater reliability. ChartSweep decreased review times by 94%–96% relative to manual review.

datasets feasible at high throughput. This is particularly important for rare diseases, diagnoses of which are often buried in plain text and not associated with a CPT code. For example, depending on the data source, between one in a million and one in 2832 patients with breast implants will be affected by BIA-ALCL.^{19–22} ChartSweep affords researchers the opportunity to review patient records using multiple search terms related to BIA-ALCL (130 minutes), including symptoms, patient demographic information, and surgery-specific key terms. It would take a manual reviewer 18-fold longer.

Previous studies have described the use of NLP and/or CPT codes to expedite the process of RCR.^{23–26} Billing codes are often entered by nonclinical administrative staff and fail to account for clinical details embedded in provider notes that are important for correct disease definition. New AI-equipped platforms are currently being developed to analyze narrative text reports and eventually assist with time-consuming RCR of large patient cohorts.²⁷

Despite significant nationwide investment in AI by healthcare organizations, few NLP tools built for medical use at one institution have successfully been repurposed for use elsewhere—their degree of technological complexity limits the ability of widespread use.²⁸ Further, access to hospital-wide billing data usually requires assistance from a back-end informatics office, which may receive hundreds of queries weekly and take extended periods of time to produce actionable datasets. We value the ability to fine-tune our queries with ChartSweep and iterate multiple times without having to involve another stakeholder. Although ChartSweep is more rudimentary than its NLP counterparts, it can be applied to any EHR platform and can be adapted for use at other institutions with relative ease. It is true to the methodological rigor of manual RCR and the scalability of nascent AI platforms.

This study should be interpreted taking into account the following limitations. Software built to automate RCR is limited to the interpretation of encoded text and cannot interpret documents scanned into a patient's medical record. Using technology to interpret photocopies of text documents is known as “bridging the semantic gap” and has not been done successfully or validated.²⁹ This means manual RCR may still be the gold standard for the review of records consisting predominantly of scanned documents. Further, ChartSweep in its current form (as tailored for use on Epic EMR at Mass General Brigham healthcare institutions) is not equipped to conduct RCR of restricted patient records. RCR of these records requires inputting a request for access and a password, which ChartSweep is not built to do. These records account for a small minority of total records and require manual inclusion.

CONCLUSIONS

Current RCR relies heavily on manual review of patient records, a technique that is time-consuming and error prone in large patient cohorts. This study describes ChartSweep, a Python-based software built to extract information from medical records, and validates its use in large unstructured datasets in the context of plastic surgery

research. ChartSweep significantly accelerates the RCR process without compromising the quality of data review and can therefore save researchers valuable time.

William G. Austen, Jr, MD

Division of Plastic and Reconstructive Surgery
Massachusetts General Hospital
15 Parkman Street, WACC 435 Boston, MA 02114
E-mail: wausten@partners.org

REFERENCES

- Gearing RE, Mian IA, Barber J, et al. A methodology for conducting retrospective chart review research in child and adolescent psychiatry. *J Can Acad Child Adolesc Psychiatry*. 2006;15:126–134.
- Gilbert EH, Lowenstein SR, Koziol-McLain J, et al. Chart reviews in emergency medicine research: where are the methods? *Ann Emerg Med*. 1996;27:305–308.
- Worster A, Haines T. Advanced statistics: understanding medical record review (MRR) studies. *Acad Emerg Med*. 2004;11:187–192.
- Alpert JS. The electronic medical record in 2016: advantages and disadvantages. *Digit Med*. 2016;2:48.
- Matt V, Matthew H. The retrospective chart review: important methodological considerations. *J Educ Eval Health Prof*. 2013;10:12.
- Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347–1358.
- Chandawarkar A, Chartier C, Kanevsky J, et al. A practical approach to artificial intelligence in plastic surgery. *Aesthet Surg J Open Forum*. 2020;2:ojaa001.
- Friedman C, Shagina L, Lussier Y, et al. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc*. 2004;11:392–402.
- Huang Y, Lowe HJ, Klein D, et al. Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon. *J Am Med Inform Assoc*. 2005;12:275–285.
- Thirukumaran CP, Zaman A, Rubery PT, et al. Natural language processing for the identification of surgical site infections in orthopaedics. *J Bone Joint Surg Am*. 2019;101:2167–2174.
- Lee SH. Natural language generation for electronic health records. *NPJ Digit Med*. 2018;1:63.
- Cohen IG, Mello MM. HIPAA and protecting health information in the 21st century. *JAMA*. 2018;320:231–232.
- Magnusson MR, Cooter RD, Rakhorst H, et al. Breast implant illness: a way forward. *Plast Reconstr Surg*. 2019;143:74S–81S.
- American Society of Plastic Surgeons. BIA-ALCL resources: patient advisory. Available at <https://www.plasticsurgery.org/patient-safety/breast-implant-safety/bia-alcl-summary/patient-advisory>. Accessed August 12, 2020.
- Smith AJ. Chart reviews made simple. *Nurs Manage*. 1996;27:33–34.
- Menger MD, Schilling MK, Schäfers HJ, et al. How to ensure the survival of the surgeon-scientist? The Homburg Program. *Langenbecks Arch Surg*. 2012;397:619–622.
- Mansukhani NA, Patti MG, Kibbe MR. Rebranding “The lab years” as “professional development” in order to redefine the modern surgeon scientist. *Ann Surg*. 2017;266:937–938.
- Davis Z, Khansa L. Evaluating the epic electronic medical record system: a dichotomy in perspectives and solution recommendations. *Health Policy and Techn*. 2016;5:65–73.
- Keech JA Jr, Creech BJ. Anaplastic T-cell lymphoma in proximity to a saline-filled breast implant. *Plast Reconstr Surg*. 1997;100:554–555.
- Brody GS, Deapen D, Taylor CR, et al. Anaplastic large cell lymphoma occurring in women with breast implants: analysis of 173 cases. *Plast Reconstr Surg*. 2015;135:695–705.

21. Becherer BE, de Boer M, Spronk PER, et al. The Dutch breast implant registry: registration of breast implant-associated anaplastic large cell lymphoma—a proof of concept. *Plast Reconstr Surg.* 2019;143:1298–1306.
22. Magnusson M, Beath K, Cooter R, et al. The epidemiology of breast implant-associated anaplastic large cell lymphoma in Australia and New Zealand confirms the highest risk for grade 4 surface breast implants. *Plast Reconstr Surg.* 2019;143:1285–1292.
23. Hanna S. *Defining Cohorts Using Current Procedural Terminology Codes in Metastatic Bone Disease: Accuracy and Implications.* Portland State University School of Public Health Poster Presentations: Portland, Oreg.;2019.
24. Mountcastle SB, Joyce AR, Sasinowski M, et al. Validation of an administrative claims coding algorithm for serious opioid overdose:a medical chart review. *Pharmacoepidemiol Drug Saf.* 2019;28:1422–1428.
25. Yerasosu D, Kwok F, Kahn JM, et al. Validation of use of billing codes for identifying telemedicine encounters in administrative data. *BMC Health Serv Res.* 2019;19:928.
26. Avoundjian T, Gidwani R, Yao D, et al. Evaluating two measures of lumbar Spine MRI overuse: administrative data versus chart review. *J Am Coll Radiol.* 2016;13:1057–1066.
27. Ananthakrishnan AN, Cai T, Savova G, et al. Improving case definition of Crohn’s disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. *Inflamm Bowel Dis.* 2013;19:1411–1420.
28. Zeng QT, Goryachev S, Weiss S, et al. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC Med Inform Decis Mak.* 2006;6:30.
29. Smith JR. The real problem of bridging the “semantic gap.” Paper presented at: International Workshop on Multimedia Content Analysis and Mining, Weihai, China, June 30-July 1, 2007.