MedCrave
Step into the World of Research

Journal of Cancer Prevention & Current Research

Research Article

Open Access

CrossMark

Electronic vs manual approaches to identify patients from the EHR for cancer clinical trials–what's feasible

Abstract

Objective: Electronic health records (EHRs) offer a platform to identify patients for clinical trials. We compared an electronic approach combining natural language processing (NLP) with query capabilities of Data Warehouse using structured and unstructured information against manual review to assess feasibility in identifying subjects for a breast cancer trial.

Materials and methods: Study included women with new metastatic, ER-positive, HER2-negative breast cancer, treated with letrozole monotherapy between January 2012 and December 2015 who did not receive prior systemic therapy for advanced disease. Concordance between approaches was assessed using Cohen's kappa statistic.

Results: 826 breast cancer cases were identified; 83 were truly metastatic, ER-positive, HER2-negative. Manual review identified 77 (93%) patients compared to 51 (61%) by NLP. Cases missed by electronic approach were due to inaccessibility of data and variability in physician documentation. Cohen's kappa was 0.36 (95% CI 0.27-0.45), indicating fair agreement. The final eligible study population included 30 women, 28 (93%) identified by manual review and 17 (57%) electronically. The electronic approach markedly reduced time spent: 44 vs. 280 hours.

Discussion: While electronic approach offers substantial cost and time savings, variability in physician documentation and inaccessibility of unstructured key data requires manual support to redress misclassification and exclusion of patients by electronic review.

Conclusion: Key common data elements need to be developed and incorporated into the clinical care process. Technological innovations are needed to lessen the pain of structured field entry. Whereas the ultimate cost savings can be substantial, there needs to be upfront investment to obtain such efficiencies.

Keywords: electronic health records, natural language processing, data warehouse, manual review, pragmatic trial, feasibility

Volume 11 Issue 4 - 2020

Nina A Bickell,[1,2] Sylvia Lin,[1] Helena L Chang,[1] Tielman Van Vleck,[3] Girish Nadkarni,[3,4] Stephen B Ellis,[3] Hannah Jacobs El,[1] Amy Tiersten,[2,4] Michael Shafir,[5] Annetine C Gelijns[1]

[1]Department of Population Health Science & Policy, Icahn School of Medicine at Mount Sinai, USA
[2]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, USA
[3]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, USA
[4]Department of Medicine, Icahn School of Medicine at Mount Sinai, USA
[5]Department of Surgery, Icahn School of Medicine at Mount Sinai, USA

Correspondence: Nina ABickell, Department of Population Health Science & Policy, Icahn School of Medicine at Mount Sinai 1 Gustave L Levy Place, Box 1077, New York, NY 10029, USA, Fax (212) 423-2998, Tel (212) 659-9567, Email nina.bikell@mssm.edu

Received: August 10, 2020 | **Published:** August 31, 2020

Background and significance

Identifying patients who are eligible for a randomized trial is often the rate-limiting step for conducting clinical trials.[1,2] Electronic Health Records (EHR), data repositories of longitudinal clinical data for large numbers of patients, may offer a more efficient approach to identify eligible study patients. Furthermore, they offer the potential to improve enrolling a diverse population of patients thereby increasing generalizability, while reducing research coordinator time to collect data that is already documented. Large groups including PCORnet Cancer Collaborative Research Group, European Get Real Consortium and American Society of Clinical Oncology (ASCO) via the minimal common oncology data elements (mCODE), are but a few of the numerous groups using EHRs to conduct pragmatic trials.[3–6] More recent studies of pragmatic trials using the EHR focus on randomization and subsequent data collection of structured variables[7–9] and do not address the challenges posed by broad types of trials, e.g., cancer, where much of needed data remain in unstructured format.

EHRs were created for clinical documentation and administrative purposes, not research.[10] Existing data is often in unstructured text and exist in a variety of locations and formats within the EHR, thus presenting a challenge to researchers. Much of the early investment in EHR pragmatic trials have focused on the low hanging fruit–trials that use structured variables and outcomes such as comparing the effectiveness of different aspirin doses on heart attacks and strokes.[6] Unfortunately, identifying potentially eligible patients and key outcomes is more challenging when needed details are typically unstructured and located in a variety of locations. Many groups, currently working on Natural Language Processing (NLP) programs seek to unlock the data secreted away in progress, pathology, radiology and scanned notes. For trials dependent on more textured and unstructured variables, numerous studies utilize varying algorithms and extraction methodologies, but few have validated NLP approaches, particularly in the cancer realm.[11]

Objective

We undertook this study to assess the utility and challenges of using a NLP approach to identify women with a new metastatic breast cancer treated with an aromatase inhibitor who could be eligible for a trial with a CDK inhibitor.[12] We use this case study to highlight varying strategies needed to maximize usability of current EHR data and suggest a variety of approaches to improve data capture.

Materials and methods

Data source

The Data Warehouse is the repository of clinical, financial and operational data sourced from over 25 data collection systems for

over 8.6 million patients' Epic EHRs. Clinical data include text based progress notes, discharge summaries, operative, pathology and radiology reports, computerized order entry of medication prescriptions, administrations and patient report, and laboratory data.

Defining the target population

The target population for our study was women, who came into the Health System between January 2012 and December 2015 with a new diagnosis of stage-IV estrogen-receptor-positive (ER-positive), human epidermal growth factor receptor 2 negative (HER2-negative) breast cancer, not previously treated and started on letrozole monotherapy as their initial systemic therapy for advanced disease.

Approach to identify study population

First, we named the key variables needed to identify the potentially eligible patient population (Table 1).

Table 1 Key variables to identify study population

	Variable	Type	Source/ Codes
1	Gender	Structured	Data Warehouse
2	Breast cancer diagnosis	Structured	Data Warehouse Problem list: ICD9 174, ICD10 C50 (excluding C50.92, male)
3	Estrogen receptor	Unstructured	Data Warehouse: Pathology report, progress notes NLP: SNOMED*416053008|Estrogen receptor positive tumor*, ER+, ER positive
4	HER2 receptor	Unstructured	Data Warehouse: Pathology report, progress notes NLP: SNOMED *705105000|Human epidermal growth factor 2 gene amplification negative, Her2 neg, Her2-*
5	New diagnosis of stage-IV/ metastatic breast cancer	Unstructured	Data Warehouse: Pathology report, progress notes, radiology reports NLP: SNOMED: *128462008|Metastases*, *55440008|M1 and 2640006|Clinical stage IV*; exclude metastasis to lymph nodes, *59441001|Structure of lymph node*
6	New letrozole treatment for new metastatic breast cancer	Structured and Unstructured	Data Warehouse Medication [orders, administration, patient report], progress notes & visit dates

NLP, natural language processing; SNOMED, systematized nomenclature of medicine

Then each data element was classified as structured or unstructured and likely sources of data specified. The initial sweep to find all potential patients for manual and electronic review searched the Data Warehouse problem lists for women with a breast cancer diagnosis and the medication lists for women treated with letrozole. This approach identified 826 women. This list was given to the groups performing manual and electronic reviews. Waiver of patient consent was approved by the IRB for this retrospective chart review(HS#16-01135).

Defining index date

The Index Date, the equivalent of a study "start" date, was the date on which letrozole was first prescribed or taken for the initial treatment of metastatic disease.

Manual approach

Manual review of records started with identifying a diagnosis of metastatic breast cancer followed by the evaluation of ER/HER2 status. We chose this strategy since identifying the presence and date of metastases at the outset would reduce the number of charts requiring review. In a few cases though, the date of metastases was identified later in the selection process. Abstractors searched pathology, radiology and progress notes for the first date of an advanced cancer diagnosis. They then identified those who were ER-positive/HER2-negative followed by letrozole therapy start date. Patients on letrozole or anastrozole in the prior year were excluded.

Electronic approach

The NLP engine used in this study was CLiX from Clinithink, LTD. The CLiX clinical NLP engine encodes textual clinical documentation as Systematized Nomenclature of Medicine (SNOMED) expressions and provides a query engine with which to identify all patients matching exact criteria. For all cases identified by the Data Warehouse, the NLP team received radiology, pathology, and progress note text reports. NLP was used on all progress notes, regardless of clinic type. These files were loaded into CLiX, which then parsed out identifiable patient facts and stored them as SNOMED expressions. A breast cancer clinician provided commonly used local vernacular (i.e. stage IV, metastatic breast cancer) to inform NLP programming language.

First, NLP identified ER and HER2 status to limit inclusion to those who were both ER-positive and HER2-negative. Queries were constructed for two related SNOMED concepts: 416053008|Estrogen receptor positive tumor and 705105000|Human epidermal growth factor 2 gene amplification negative. Unlike manual review, receptor status was identified first rather than presence of metastatic disease as we were forewarned of the challenge of interpreting dates with this software. Next, to identify the diagnosis of breast cancer metastasis, queries were constructed to identify patients with the presence of the following three SNOMED concepts: 128462008|Metastases, 55440008|M1 and 2640006|Clinical stage IV. To determine whether the SNOMED codes correctly identified patients, the study team reviewed randomly selected cases to manually determine whether

those electronically classified as metastatic were truly metastatic. This review identified a number of reasons why the terms were too broad and metastasis queries were refined to exclude metastasis to lymph nodes, 59441001|Structure of lymph node, and educational EHR template language that described counseling about metastatic disease included in charts of some patients with regional spread.

Identifying date of index visit

The next step was to identify the index visit in which a metastatic diagnosis was first documented and a prescription for letrozole ordered or evidence in the notes of the day patient started taking letrozole. Because some women might be started on treatment prior to advanced cancer diagnosis based on clinical suspicion alone, and because not all patients undergo biopsy to confirm metastasis, we applied an algorithm of a new letrozole prescription ±90 days of the metastatic date, provided no letrozole had been prescribed in the prior year. Patients treated with either letrozole or anastrozole during the year prior were excluded. As progress notes rarely reference a date of metastasis, we assigned the date of the first visit note identifying a metastasis as the "metastatic date." This date was forwarded to the Data Warehouse. The Data Warehouse created a ±90 day window around the first metastatic date and designated the first letrozole treatment within that window as the index date. The ±90 day window was not used by manual reviewers because they could more easily identify the onset of metastasis and letrozole treatment. The letrozole records were based on prescriptions and medications administered and did not include patient reported medications.

Comparing electronic to manual identification of patients

This algorithm identified potentially eligible, newly metastatic, ER-positive, HER2-negative breast cancer patients on first line letrozole therapy. We then compared the patients identified electronically with those identified manually to determine the overlap and differences and to understand the sources of these differences. Errors in identifying the target population (first step) were carried forward to the identification of the index visit date (second step). If a case was missed by an approach (electronic or manual) in the first step, it was not available for analysis in the second step. If a case was wrongly selected in the initial step, it was still evaluated in the subsequent step. Therefore, while we reviewed all errors in the first stage, we focused only on new errors in the second stage. The new errors consist of those patients who were rightly chosen by both approaches in the first step but incorrectly selected or incorrectly excluded in the second step.

Results

Identifying the target population

The Data Warehouse identified 826 women with an ICD-9-CM/ICD-10-CM code for breast cancer and treatment with letrozole between January 2012 to December 2015, of whom 83 were true metastatic, ER-positive, HER2-negative cases. Manual review correctly identified 77 (93%) patients compared to 51 (61%) by NLP. In total, the manual approach selected 83 women who had ER-positive, HER2-negative metastatic breast cancer but 6 were false positives (Figure 1), while NLP selected 122 cases with 71 false positives. Of the 83 true cases, 45 were identified by both manual review and NLP. The 32 cases missed by NLP were primarily due to the absence of records (21/32; 66%) provided by the Data Warehouse (Table 2).

This included inaccessibility of non-machine-readable documents (scanned jpeg, pdf files), operative notes, and pathology reports from visits that appeared as an invalid/non-standard visit number in the Data Warehouse. However, such visits existed in Epic and were viewable by the manual reviewers. The remaining missed cases (11/32; 34%) were due to variations in medical doctor (MD) documentation of "ER-positive, HER2-negative" that were not understood by NLP–for example, "ER-positive" versus "positive for ER" or "HER2-negative" versus "HER negative".

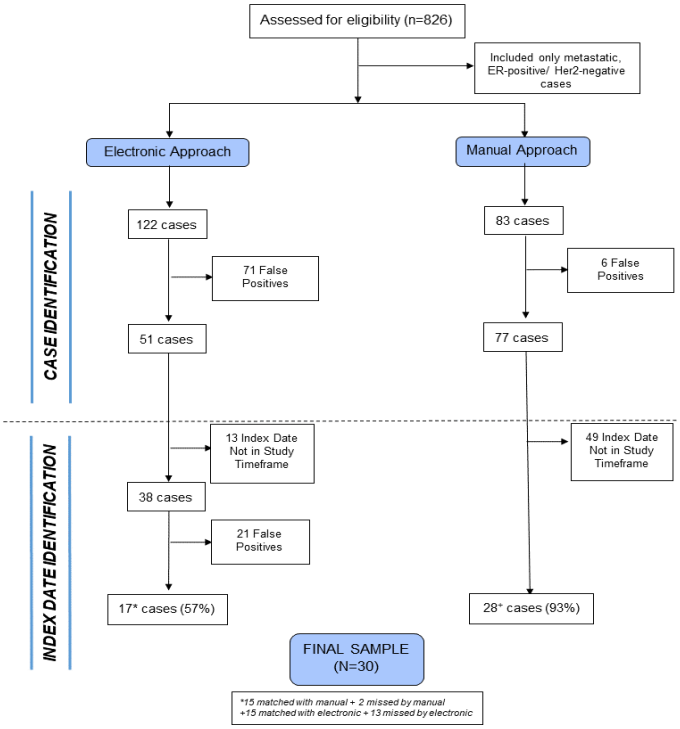

Figure 1 Consort diagram.

Citation: Bickell NA, Lin S, Chang HL, et al. Electronic vs manual approaches to identify patients from the EHR for cancer clinical trials–what's feasible. *J Cancer Prev Curr Res.* 2020;11(4):99–104. DOI: 10.15406/jcpcr.2020.11.00436

Table 2 Causes of & potential solutions to electronic-manual case identification errors

	Defining the target population	**Current challenges**	**Types**	**Potential solutions**
1	Electronic	Variation in MD documentation and NLP programming omission of variation in documentation	• MD vernacular • Not stage-IV (i.e. metastatic to lymph node= regional disease) • Receptor status on contralateral breast	• Create structured variables for histology, receptor status, recurrence, progression, death • Generate templates for MD to minimize variability • Transfer learning method • Ensure adequate sample of specialists for language specification
		Not machine readable	• Scanned documents (i.e. jpeg, pdf files)	• Optical Character Recognition (OCR)
		No electronic data available	• Surgery notes • Lack of shared information (outside hospital)	• Expand pool of data sources • Identify the parameters of electronic data sources
		Data transcription issue	• Invalid/ non-standard visitnumber in Data Warehouse	• Technical solution required
2	Manual	Human Error	• Missed documented metastatic stage • Differential application to assess eligibility (primary vs metastatic biopsy) • Misclassification error	• Clarify parameters of study sample • Apply same case ascertainment criteria • Abstractor training
	Identifying the study start date	**Current challenges**	**Types**	**Potential solutions**
1	Electronic	Programming specifications for study design	• Stage-IV prior to study time frame or initial visit	• Refine NLP algorithms to correctly identify 1st date of metastatic diagnosis
		Differential application to assess eligibility	• Consult visits • Consumption of letrozole vs order of letrozole • Clinical vs objective dx • ±90 day window applied to electronic approach only	• Expand window to include data outside study time frame • Exclude if patient comes in for 1 visit
		Variation in MD documentation and NLP programming omission of variation in documentation	• Not stage-IV (i.e. metastatic to lymph node	• Create structured variables for histology, receptor status, recurrence, progression, death • Generate templates for MD to minimize variability • Transfer learning method • Ensure adequate sample of specialists for language specification
		Not detected in	• Unknown reason	• Technical solution required

MD, medical doctor

Variations in MD documentation also contributed largely to the false positives incurred by NLP (69/71; 97%), mainly because of the way "metastasis" was used. When multiple cancers were present, it was difficult to linguistically correlate the body location of an identified metastasis with the cancer with which it was associated (i.e. metastatic to the lymph nodes). As such, a decision had to be made whether to look only at instances where a location could be identified, reducing false positives (prioritizing specificity) or look without a location to ensure identifying all patients (prioritizing sensitivity). Going without a body location, 60 cases selected by NLP did not have stage-IV breast cancer. If an identifiable body location was required, false positives for stage-IV fell to 9 while increasing false negatives by an additional 5 patients. Deeming the risk of missing a critical patient to be greater than the cost of additional manual review, sensitivity was prioritized.

Manual review was not without error. Six manually identified cases were misclassified as metastatic, ER-positive, HER2- negative (e.g., incorrectly classifying triple negative breast cancer as ER-positive). Notably, another 6 were missed by manual review but picked up by NLP. Of these, half had insufficient data on receptor status for their metastatic lesion. NLP, instead, picked up receptor status of the primary tumor which was confirmed acceptable by the sponsor later on.

Cohen's kappa for assessing the level of concordance in case identification between manual review and NLP was 0.36 (95% CI, 0.27-0.45), suggesting fair agreement between the 2 approaches.

Identifying the date of the index visit

Next, the algorithm identified the index date for each patient and assessed letrozole/anastrozole treatment in the year prior to the index date so women who had been on these treatments would be excluded (Figure 1). On the electronic side, identification of the index date and further exclusions were done by the Data Warehouse upon receipt of the first date of metastasis from NLP. After the letrozole/anastrozole exclusion criteria were applied, 28 women remained with the manual approach and 38 with the electronic approach, of whom 21 were false positives by the latter. Fifteen patients were chosen by both approaches.

Thus, among the 826 women originally identified as having breast cancer and a letrozole prescription between years 2012- 2015, a total of 30 eligible metastatic, ER-positive, HER2-negative patients who started letrozole as the first line therapy were identified by manual and electronic approaches. The manual approach correctly identified 28 (93%) of the 30 eligible patients while the electronic approach identified 17 (57%). Seven (54%) of the 13 eligible women missed by the electronic approach and 6 (29%) of its 21 false positives came from the 45 patients selected by both manual and NLP in the first step. This begs the question why the 7 women weren't chosen by the Data Warehouse in the second step, opting instead for 6 ineligible ones.

Closer reexamination (Table 2) showed that the vast majority were due to differences in programming specifications and differential applications to assess eligibility.

Manually, eligible cases were marked as ineligible because patients came in for a consult visit or were prescribed letrozole, but never took the medication. The electronic approach included patients for whom letrozole was ordered, sometimes mistakenly equating computerized order entry as taking the drug. Electronically, some ineligible cases were included because initial visit notes indicated that patient had metastatic breast cancer, but were actually diagnosed prior to study time frame. Date of diagnosis for some cases was identified electronically based on a clinical diagnosis, whereas manually, it was based on an objective tissue diagnosis. Letrozole was not prescribed until confirmation of an objective tissue diagnosis, affecting the Data Warehouse's ability to find a prescription within the ±90 day window. Finally, manual review did not incur further errors; the 2 excluded eligible patients were missed in the first step.

Feasibility

On average, manual abstraction took 20 minutes per chart to identify cases (N=826) and 10 minutes per chart (N=28) to identify index date, totaling 280 hours. NLP document loading, query construction, running, refinement and re-running took 4 hours; the Data Warehouse spent 40 hours creating queries, running and re-running algorithms for a total of 44 hours for the electronic approach to identify patients. An additional 100 hours was spent reviewing and reconciling the differences between the manual and electronic approaches to identify our final eligible sample. Given time constraints, the comparison of electronic vs manual was limited to case identification and did not include variable identification since many of the variables required dates, a known challenge for the NLP program. The electronic approach markedly reduced the time spent on case identification: 44 vs. 280 hours.

Discussion

Our study assessing the feasibility of using the EHR for pragmatic clinical trials identified some key strengths, limitations and critical take home messages to help guide process improvement for future use. The EHR presents a potent platform to identify potential patients for clinical trials. To identify eligible patients, 6 key variables were needed, two of which were available onlyas unstructured data. The electronic approach took a total of 44 hours to review 826 potential cases, identify 57% of the eligible patients, and refine queries and algorithm.

For highly prevalent conditions, and certainly, for conditions utilizing only structured variables, an electronic approach to identify cases is practical. However, for studies requiring unstructured data to identify relevant cohort, electronic review remains challenging due to variability in physicians' documentation. Difficulties arose as notes frequently referenced a metastasis several sentences away from the section describing the carcinoma. Acronym expansion was also challenging. While "Mets" was frequently used to refer to metastases, it was used occasionally to refer to other things like "metabolic equivalents". Challenges of laterality arose for women with >1 breast cancer. The biggest source of error was due to variation in MD documentation and NLP programming omission of variation in documentation. Common causes of error in case identification could be avoided by creating and implementing key structured data fields. As cancer clinical trials move towards greater precision medicine, they require specific pathologic and genomic details including stage, receptor status, genomic mutation status, recurrence, and disease progression. These variables could easily be structured and, in fact, pathology software packages now include structured fields and ASCO is leading efforts of mCODE,[13,14] an approach that can address the majority of mismatch causes.

Another source of error was due to the absence of records because data was located in notes such as media and surgery that were not analyzed by NLP as we anticipated the yield from such sources to be

low. Implementing an Optical Character Recognition (OCR) software can enable the decoding of images of typed, written, or handwritten text into electronically recognizable text to address non-machine-readable documents.[15] For a small number of cases, visit notes were not coded in a standard way so the Data Warehouse did not transfer those notes to NLP. Data Warehouse staff were unaware that such visit codes were used and thus, unable to preemptively address this issue. A technical solution is required to address data transcription issues.

An unforeseen challenge was related to unanticipated programming specifications for study design and differential application to assess eligibility from the outset. These could have been minimized by clearly delineating additional exclusions and specifications a priori. For example, patients with only 1 visit could have been excluded, since their start of treatment was unknown. However, known limitations of the EHR (e.g., prescription of drug does not equate to consumption) and Clinithink identification of dates could not be avoided. Our study is somewhat limited in that we focused on a single breast cancer trial. While limited to specific cancer type, the challenges inherent in utilizing unstructured variables is common to many clinical trials.

Even as we move towards structuring these common data elements, we will still be faced with the challenge of data completeness and burden of data entry. Who will enter these Common Data Elements? Physicians? That scenario is unlikely as physician burnout attributed to the EHR is great.[13] Requiring more data entry runs counter to the move to reduce strokes and clicks and attend to physician wellness.[16] In redesigning clinical trials infrastructure, some of the costs saved with electronic case identification can be applied to train research coordinators to complete data entry.

Artificial intelligence with pattern recognition using deep neural networks has been used effectively in radiology and such approaches may be useful to identify less complex cases, like patients with an abdominal aneurysm >5cm.[17] However, the efficacy of this innovative approach to identify patients with more complicated conditions, eligibility requirements and restrictions remains unknown.

Conclusion

The EHR presents a viable and potentially time saving approach for pragmatic trials. However, key common data elements for unstructured fields need to be developed and then incorporated into clinical care documentation processes. Technological innovations are needed to lessen the pain of structured data field entry. Whereas the ultimate cost savings can be substantial, there needs to be upfront investment to obtain such efficiencies.

Acknowledgments

We would like to acknowledge Jack Mardekian, PhD from Pfizer who assisted in clarifying the variables of interest.This work was presented at Academy Health 2019.

Declaration of conflicting interests

Dr. Amy Tiersten serves as a consultant for Immunomedics, AstraZeneca, Novartis and Eisai, and gets research funding from Pfizer, Novartis, Genentech, Lilly and AstraZeneca.

Dr. Van VleckTielman was part of launching Clinithink and retains a financial interest in the company. All other authors have no conflict of interest to report.

Funding

This work was supported by Pfizer.

References

1. Stewart DJ, Batist G, Kantarjian HM, et al. The urgent need for clinical research reform to permit faster, less expensive access to new therapies for lethal diseases. *Clin Cancer Res*. 2015;21(20):4561–4568.
2. Stewart DJ, Whitney SN, Kurzrock R. Equipoise lost: ethics, costs, and the regulation of cancer clinical research. *J Clin Oncol*. 2010;28(17):2925–2935.
3. Clinical Trials Transformation Initiative. 2017.
4. MITRE Announces Compass TM, a New Open-Source Application to Collect Common Oncology data. 2019.
5. Zuidgeesta MGP, Goetzb I, Groenwold RHH, et al. Series: Pragmatic trials and real world evidence: Paper 1. Introduction. *J Clin Epidemiol*. 2017;88:7–13.
6. The National Patient-Centered Clinical Research Network. 2018.
7. Simon KC, Tideman S, Hillman L, et al. Design and implementation of pragmatic clinical trials using the electronic medical record and an adaptive design. *JAMIA Open*. 2018;1(1):99–106.
8. Hills TA, Beasley R. Pragmatic randomised clinical trials using electronic health records: general practitioner views on a model of a priori consent. *Trials*. 2018;19(1):278–280.
9. Mc Cord KA, Ewald H, Ladanie A, et al. Current use and costs of electronic health records for clinical trial research: a descriptive study. *CMAJ Open*. 2019;7(1):E23–E32.
10. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research. *Med Care*. 2013;51(8 Suppl 3):S30–S37.
11. Ford E, Carroll JA, Smith HE, et al. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc*. 2016;23(5):1007–1015.
12. Finn RS, Martin M, Rugo HS, et al. Palbociclib and Letrozole in Advanced Breast Cancer. *N Engl J Med*. 2017; 375(20):1925–1936.
13. Casati B, Bjugn R. Structured Electronic Template for Histopathology Reporting on Colorectal Carcinoma Resections: Five-Year Follow-up Shows Sustainable Long-Term Quality Improvement. *Arch Pathol Lab Med*. 2012;136(6):652–656.
14. mCODE: Creating a Set of Standard Data Elements for Oncology EHRs. 2020.
15. Optical Character Recognition (OCR). 2019.
16. Collier R. Rethinking EHR Interfaces to Reduce Click Fatigue and Physician Burnout. *CMAJ*. 2018;190(33):E9994–E9995.
17. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56.