Temporal Trends in Clinical Evidence of 5-Year Survival Within Electronic Health Records Among Patients With Early-Stage Colon Cancer Managed With Laparoscopy-Assisted Colectomy vs Open Colectomy

Jue Hou, PhD; Rachel Zhao, BSc; Tianrun Cai, MD, PhD; Brett Beaulieu-Jones, PhD; Thany Seyok, MS; Kumar Dahal, MS; Qianyu Yuan, PhD; Xin Xiong, BSc; Clara-Lea Bonzel, MSc; Claire Fox, MPH; David C. Christiani, MD; Thomas Jemielita, PhD; Katherine P. Liao, MD, MPH; Kai-Li Liaw, PhD; Tianxi Cai, ScD

Abstract

IMPORTANCE Temporal shifts in clinical knowledge and practice need to be adjusted for in treatment outcome assessment in clinical evidence.

OBJECTIVE To use electronic health record (EHR) data to (1) assess the temporal trends in treatment decisions and patient outcomes and (2) emulate a randomized clinical trial (RCT) using EHR data with proper adjustment for temporal trends.

DESIGN, SETTING, AND PARTICIPANTS The Clinical Outcomes of Surgical Therapy (COST) Study Group Trial assessing overall survival of patients with stages I to III early-stage colon cancer was chosen as the target trial. The RCT was emulated using EHR data of patients from a single health care system cohort who underwent colectomy for early-stage colon cancer from January 1, 2006, to December 31, 2017, and were followed up to January 1, 2020, from Mass General Brigham. Analyses were conducted from December 2, 2019, to January 24, 2022.

EXPOSURES Laparoscopy-assisted colectomy (LAC) vs open colectomy (OC).

MAIN OUTCOMES AND MEASURES The primary outcome was 5-year overall survival. To address confounding in the emulation, pretreatment variables were selected and adjusted. The temporal trends were adjusted by stratification of the calendar year when the colectomies were performed with cotraining across strata.

RESULTS A total of 943 patients met key RCT eligibility criteria in the EHR emulation cohort, including 518 undergoing LAC (median age, 63 [range, 20-95] years; 268 [52%] women; 121 [23%] with stage I, 165 [32%] with stage II, and 232 [45%] with stage III cancer; 32 [6%] with colon adhesion; 278 [54%] with right-sided colon cancer; 183 [36%] with left-sided colon cancer; and 222 [43%] with sigmoid colon cancer) and 425 undergoing OC (median age, 65 [range, 28-99] years; 223 [52%] women; 61 [14%] with stage I, 153 [36%] with stage II, and 211 [50%] with stage III cancer; 39 [9%] with colon adhesion; 202 [47%] with right-sided colon cancer; 39 [9%] with left-sided colon cancer; and 201 [47%] with sigmoid colon cancer). Tests for temporal trends in treatment assignment ($\chi^2 = 60.3; P < .001$) and overall survival ($\chi^2 = 137.2; P < .001$) were significant. The adjusted EHR emulation reached the same conclusion as the RCT: LAC is not inferior to OC in overall survival rate with risk difference at 5 years of $-0.007$ (95% CI, $-0.070$ to $0.057$). The results were consistent for stratified analysis within each temporal period.

CONCLUSIONS AND RELEVANCE These findings suggest that confounding bias from temporal trends should be considered when conducting clinical evidence studies with long time spans.

Key Points

Question How might investigators assess and adjust for temporal trends appropriately to correct for confounding bias in studies using clinical evidence?

Findings This comparative effectiveness study of 943 patients with early-stage colon cancer detected substantial temporal trends by emulating the Clinical Outcomes of Surgical Therapy (COST) Study Group Trial using electronic health records data from 2006 to 2017. Through stratification by calendar time and cotraining of time-varying models, this emulation reached a conclusion agreeing with the COST Study Group Trial.

Meaning These findings suggest that with proper adjustment of confounding bias from temporal trends, clinical studies conducted with data covering a long time span could supplement randomized clinical trials in the assessment of treatment outcome and its evolution over time.

Open Access. This is an open access article distributed under the terms of the CC-BY-NC-ND License.


June 23, 2022 1/13

Downloaded From: https://jamanetwork.com/ on 06/25/2022
Abstract (continued)

Stratification of calendar time and cotraining of models is one solution. With proper adjustment, clinical evidence may supplement RCTs in the assessment of treatment outcome over time.


Introduction

Electronic health record (EHR) data, with detailed longitudinal clinical information, have the potential to generate clinical evidence to assess treatment outcomes. Recently, regulatory agencies have created a framework with guidance on approaches for leveraging clinical evidence to evaluate the effectiveness of medical products.1 This, together with the wider availability of EHR data, has led to an increasing number of studies comparing EHR-based analyses with randomized clinical trials (RCTs) in recent years, including the use of external controls,2-4 supporting a new indication for approved therapy5 and pragmatic trial designs.6

In clinical practice, secular trendsin management and treatment options can lead to changes in the effectiveness of a given treatment. Although the treatment remains the same, other factors may alter the treatment response or outcome. For example, the overall survival of patients with early-stage cancer after surgery has improved over time as targeted therapies have become available for recurrence and metastasis. Temporal trends are listed as considerations for using EHR external control for RCTs.7 Moreover, these temporal trends may cause automated analytical tools trained in 1 year to fail at a later point in time.8

Properly understanding and accounting for temporal trends is a crucial step when comparing treatment effectiveness using longitudinal EHR data across calendar years. Shifts in patient population characteristics, treatment patterns, and responses all have significant implications for medical practice and policy making. When both treatment and response patterns shift over time, the calendar year becomes a confounding factor. Reporting of the association between treatment and response should also reflect its time-dependent property. On the other hand, it is important to strike a balance between the temporal adjustment and limited sample size within each calendar time frame. One potential strategy is to adopt a cotraining strategy for models across different periods by allowing some associations identified from data-driven selection to change over time. The objective of this study was to investigate temporal trends and correct for their confounding bias in an emulation of a completed RCT using EHR data. The result is compared with that of the RCT.

Methods

The Mass General Brigham Institutional Review Board approved the use of the EHR data for this study with a waiver of informed consent owing to use of deidentified data. This study followed the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) reporting guideline. eFigure 1 in the Supplement illustrates the study schematics.

The RCT

The multi-institutional Clinical Outcomes of Surgical Therapy (COST) Group Study RCT on laparoscopic-assisted colectomy (LAC) vs open colectomy (OC), conducted from August 1994 to August 2001, was chosen as the target of emulation9 and was reported by Nelson et al in 2004.10 The overall finding of this trial was the noninferiority of LAC compared with OC. The individual-level data, study protocol, and report sheet were made available through Project Data Sphere.11 We extracted data regarding the type of surgical procedure, follow-up duration, vital status, and baseline features, including age, sex, cancer clinical stage, presence of colon adhesion, and tumor location. The tumor location was categorized into the left, right, or sigmoid segment of the colon. Inclusion
criteria consisted of a clinical diagnosis of adenocarcinoma of the colon (histologic confirmation was required at surgery), being 18 years or older, and the absence of prohibitive abdominal adhesions. Exclusion criteria included advanced local or metastatic disease, rectal or transverse colon cancer, acute bowel obstruction or perforation due to cancer, and severe medical illness. Inflammatory bowel disease, familial polyposis, pregnancy, or concurrent or previous malignant tumor also precluded enrollment. Patients were required to have a recent radiological test (42 days within registration) and undergo a colectomy within 21 days, so we interpreted the requirement as the implicit eligibility that a colectomy must be performed within 90 days after the colorectal cancer diagnosis. The list of key eligibility criteria, along with the EHR variables corresponding to the RCT eligibility criteria and other features and the extraction rules, are provided in eTable 1 in the Supplement.

**EHR Cohort for Emulating the RCT**

We created a data mart from the Mass General Brigham Research Patient Data Registry consisting of patients with at least 1 International Classification of Diseases, Ninth Revision, or International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, code for colon cancer by January 1, 2020. We further classified the colon cancer status for patients in the data mart based on the unsupervised multimodal automated phenotyping algorithm validated by 171 labels. Identified patients with colon cancer who underwent a colectomy formed the EHR LAC and OC groups. We identified the first colectomy according to the Current Procedure Terminology code (eTable 2 in the Supplement). The date of the first colectomy code after the colon cancer diagnosis was used as the baseline date for each patient. We included only patients whose colectomy date was 2006 or later, when LAC became widely available. We also excluded patients who underwent a colectomy after January 1, 2018, owing to their short follow-up. Patients with multiple surgical codes around the first Current Procedure Terminology code for colectomy (3 days before to 3 days after) were excluded because they likely underwent more complex procedures. Patients whose colectomy dates were 90 days after their colorectal cancer diagnosis were also excluded to remain consistent with the RCT eligibility criteria. We extracted the EHR features that can be mapped to the target RCT eligibility criteria and features (eTable 2 in the Supplement).

Key eligibility criteria of the RCT were matched in our emulation. We excluded patients with (1) diagnosis codes of prior cancers, adhesions, Crohn disease, familial polyposis, and chronic ulcerative colitis; (2) recent diagnosis codes of colon obstruction and perforation; (3) missing stage or stage IV extraction from the natural language–processing interpreter for cancer extraction algorithm; and (4) tumor location outside the right-sided, left-sided, and sigmoid colon or multiple tumor locations. We refer to the final subcohort treated with LAC or OC and satisfying the eligibility criteria as the EHR emulation cohort.

Additional features not considered by the RCT were extracted for confounding adjustments. We extracted obesity and additional health care features before the colectomy, including use of health care resources and existing comorbidities as defined by International Classification of Disease codes grouped into clinically relevant groups. To assess temporal trends, we categorized patients into 3 groups according to the year of colectomy: 2006 to 2009, 2010 to 2013, and 2014 to 2017. We refer to these intervals as temporal periods. We split the 12 years into 3 periods to allow for multiple change points while maintaining a reasonable sample size within each temporal period to ensure stable estimation. Details on EHR cohort construction and data curation are provided in eMethods 1 in the Supplement.

**Statistical Analysis**

Data were analyzed from December 2, 2019, to January 24, 2022. We assessed the temporal trends of the patient clinical profile, the propensity for selecting LAC vs OC, and the patient outcome. Aiming to emulate the RCT, we defined the average treatment effect parameter as the difference between LAC vs OC in 5-year overall survival rate for the EHR-derived target patient populations.
constructed to match the RCT patient characteristics on the key risk factors, including sex, cancer clinical stage, tumor location, and colon adhesion. Overall survival was the primary end point for the COST Group Study RCT, has been the criterion standard measure for cancer trials,\textsuperscript{17} and captures surgery-related mortality,\textsuperscript{18} whereas other end points such as progression or cancer-specific mortality are also not readily available in EHRs. To account for confounding, we adjusted for clinically relevant variables, including age, sex, cancer stage, tumor location, colon adhesion, procedure subtypes, and obesity, as well as a broad range of other comorbidities. We adopted a doubly robust causal modeling strategy\textsuperscript{19} that combines the regression adjustment approach via outcome regression and the propensity score (PS) weighting approach. To account for temporal changes, we allowed the covariate associations in both the outcome regression and PS models to vary across the temporal periods but adopted a data-driven cotraining strategy to select temporal trends as well as confounding factors.\textsuperscript{20-22} We assessed the association of temporal trends with outcome regression by testing whether the coefficients of the confounders on the mortality risk differ across temporal periods. Similarly, we tested whether the odds ratios of the confounders on treatment assignment in the PS model differed by temporal periods. To further illustrate the association of temporal trends as a confounding factor, we additionally performed causal modeling using the entire EHR emulation cohort without adjusting for temporal periods as a benchmark analysis. R, version 4.1.1 was used to conduct analyses. Tests were 2-tailed (ie, to test for difference rather than superiority/inferiority).

To form EHR-derived target populations that mimic the RCT population, we considered 4 EHR source populations according to time of colectomy: 2006 to 2017, 2006 to 2009, 2010 to 2013, and 2014 to 2017. For each EHR source population, we created the corresponding EHR-derived target population by weighting patients in the source population to match the key risk factors of the RCT population and calculate the average treatment effect. We constructed SEs and CIs based on the bootstrap method with 10 000 replications. Details of the statistical analyses are provided in eMethods 2 in the Supplement.

Results

Study Cohorts
A total of 943 patients were included in the EHR emulation cohort (452 [48%] men and 491 [52%] women; median age, 63 [range, 20-99] years) (Figure 1 and eMethods 1 in the Supplement). Of these, 518 patients underwent LAC (median age, 63 [range, 20-95] years; 250 [48%] men and 268 [52%] women; 121 [23%] with stage I, 165 [32%] with stage II, and 232 [45%] with stage III cancer; 32 [6%] with colon adhesion; 278 [54%] with right-sided colon cancer; 18 [3%] with left-sided colon cancer; and 222 [43%] with sigmoid colon cancer). A total of 425 patients underwent OC (median age, 65 [range, 28-99] years; 202 [47%] men and 223 [52%] women; 61 [14%] with stage I, 153 [36%] with stage II, and 211 [50%] with stage III cancer; 39 [9%] with colon adhesion; 202 [47%] with right-sided colon cancer; 39 [9%] with left-sided colon cancer; and 201 [47%] with sigmoid colon cancer). The baseline variables of the EHR cohort are presented in Table 1 alongside those of the target RCT cohort. The population sizes of the 4-year temporal periods were 231 patients for 2006 to 2009 (81 undergoing LAC and 150 undergoing OC), 277 for 2010 to 2013 (132 undergoing LAC and 145 undergoing OC), and 435 for 2014 to 2017 (305 undergoing LAC and 130 undergoing OC) (see eFigure 1-4 in the Supplement for a summary of multimodal automated phenotyping, extraction of body mass index, and follow-up durations).

Compared with the RCT cohort, the EHR emulation cohort differed most among patients with stage I colorectal cancer. Overall, there was a higher prevalence overall of stage I cancer in the RCT (321 of 839 [38%]) compared with the EHR emulation cohort (182 of 943 [19%]). In the RCT cohort, 425 patients (51%) underwent LAC and 414 (49%) underwent OC, whereas in the EHR emulation cohort, 518 of 943 (55%) underwent LAC and 425 (45%) underwent OC (Table 1). In addition, colon adhesion prevalence was higher in the RCT cohorts (151 of 839 [18%]) compared with the EHR
emulation cohorts (71 of 943 [7%]). Meanwhile, the prevalence of sigmoid colon cancer was lower in the RCT cohorts (320 of 839 [38%]) compared with the EHR emulation cohorts (423 of 943 [45%]).

**Temporal Trends**

The number of patients who fulfilled the eligibility criteria increased during the temporal periods, with 231 in 2006 to 2009, 277 in 2010 to 2013, and 435 in 2014 to 2017 (Figure 1). Eligible patients increasingly underwent LAC, from 81 (35%) in 2006 to 2009 through 132 (48%) in 2010 to 2013 and 305 (70%) in 2014 to 2017 (Figure 1).

**Figure 2** summarizes the temporal trends in the distribution of the confounders, the association of the confounders with treatment assignment, and the association with overall survival. Interestingly, the data-driven analysis identified temporal trends in clinically relevant variables (eg, cancer clinical stage) along with variables seemingly not associated with colectomy (eg, codes for sciatica) (eResults in the Supplement). The temporal adjustments in both PS and outcome regression models explained significant variation of LAC assignment and overall survival ($\chi^2$ test for PS model, 60.3 [$P < .001$]; $\chi^2$ test for outcome regression model, 137.2 [$P < .001$]).

**EHR Emulation: LAC vs OC**

After controlling for confounding variables as well as temporal shift, we observed that LAC was not associated with a significant change in overall survival, consistent with the RCT findings. The estimated mean survival rates for the RCT target population when treated with LAC or OC using the entire EHR emulation cohort demonstrated similar trajectories and strongly overlapping confidence bands for pooled and time-stratified analyses (Figure 3). The estimated average treatment effect
(LAC vs OC) based on the full EHR emulation cohort was −0.007 (95% CI, −0.070 to 0.057), suggesting that 5-year survival was similar after LAC and OC (Table 2). Using time-stratified EHR emulation cohorts as target populations resulted in similar conclusions.

We report the comparison with benchmark analyses in the eResults in the Supplement. The contrast between the benchmark analyses and the temporally adjusted analysis demonstrated the potential confounding bias due to ignoring temporal trends (eFigure 5 in the Supplement).

Although the estimated treatment differences based on EHR cohorts were consistent with those reported in the RCT cohorts, the estimated overall survival rates were higher in the EHR emulation cohorts than in the RCT cohorts (average treatment effect, −0.013 [95% CI, −0.077 to 0.093]) (Table 2 and Figure 3). Compared with the RCT conducted from 1995 to 2001, the estimated 5-year survival rates based on the EHR emulation analyses were 0.074 higher (95% CI, 0.011-0.131) for the OC group and 0.061 higher (95% CI, −0.001 to 0.122) for the LAC group. The estimated overall survival rates also increased during the study period in the EHR emulation cohort. The estimated 5-year survival rates in the time-stratified analyses increased from 0.767 (95% CI, 0.694-0.837) for the OC group and 0.819 (95% CI, 0.705-0.922) for the LAC group in 2006 to 2009 through 0.855 (95% CI, 0.791-0.907) for the OC group and 0.834 (95% CI, 0.767-0.905) for the LAC group in 2010 to 2013 and 0.885 (95% CI, 0.803-0.948) for the OC group and 0.857 (0.771 to 0.923) for the LAC group in 2014 to 2017.

**Discussion**

In this comparative effectiveness study, we demonstrate the association of temporal trends with study outcomes using EHR data by comparing results of an EHR emulation cohort with or without temporal adjustments with RCT results. Further, we showcased approaches that enable us to account for the temporal trends in clinical evidence through stratification by calendar time and cotraining of the time-dependent models. After accounting for confounding in temporal trends, our EHR emulation replicated the RCT findings of noninferiority in LAC compared with OC for overall survival within 5 years. We identified various temporal trends in the analysis, including the shift in patient population, treatment pattern, and overall survival rate.

**Table 1. Characteristics of the Treatment Groups**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Target RCT cohort</th>
<th>RCT eligible in EHR cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment group</td>
<td>Treatment group</td>
</tr>
<tr>
<td></td>
<td>LAC (n = 425)</td>
<td>LAC (n = 518)</td>
</tr>
<tr>
<td></td>
<td>OC (n = 414)</td>
<td>OC (n = 425)</td>
</tr>
<tr>
<td>Age, median (range), yb</td>
<td>69 (29-94)</td>
<td>63 (20-95)</td>
</tr>
<tr>
<td></td>
<td>70 (28-96)</td>
<td>65 (28-99)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>217 (51)</td>
<td>250 (48)</td>
</tr>
<tr>
<td>Women</td>
<td>208 (49)</td>
<td>268 (52)</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>176 (41)</td>
<td>121 (23)</td>
</tr>
<tr>
<td>II</td>
<td>136 (32)</td>
<td>165 (32)</td>
</tr>
<tr>
<td>III</td>
<td>113 (26)</td>
<td>232 (45)</td>
</tr>
<tr>
<td>Colon adhesion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon adhesion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right colon cancer</td>
<td>94 (22)</td>
<td>32 (6)</td>
</tr>
<tr>
<td>Left colon cancer</td>
<td>32 (7)</td>
<td>18 (3)</td>
</tr>
<tr>
<td>Sigmoid colon cancer</td>
<td>162 (38)</td>
<td>222 (43)</td>
</tr>
<tr>
<td>Abbreviations: EHR, electronic health record; LAC, laparoscopy-assisted colectomy; NA, not applicable; OC, open colectomy; OR, odds ratio; RCT, randomized clinical trial.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Except for age, all characteristics are binary/categorical, and the summary statistics are expressed as No. (%) within the group. Percentages have been rounded and may not total 100.

b As reported by the target RCT.
Patients enrolled in RCTs may differ significantly from the general patient population in health care systems with regard to the distribution of their risk profiles and disease characteristics. Therefore, the efficacy reported from RCTs may not be always replicable in health care systems.\textsuperscript{23}

![Figure 2. Model Coefficients in Propensity Score (PS) and Cox Proportional Hazards Models](image-url)

Box sizes reflect the mean of the variable in the period relative to the overall mean. Colors indicate the effect size in the PS and Cox models from negative to positive. Side-by-side boxes in the Cox model panel indicate different associations with OC and LAC through interaction of the variable with treatment assignment. Missing square indicates the variable is not selected for the model.
Furthermore, changes in overall disease management may not be observed or foreseen in RCTs but can be assessed using EHR data and provide insights for clinicians when making treatment decisions. Factors such as increasing trends of screening, improvement of surgical skills, or better previous lines of treatment that are associated with the treatment outcome may change over time, resulting in changes in effectiveness.\textsuperscript{24,25}

The trend of increasing adoption of LAC over time in the EHR emulation cohort was also observed in US nationwide data.\textsuperscript{26-28} The increasing trend of patients with stages II and III cancer assigned to LAC could be associated with the accumulation of evidence from clinical trials suggesting

Figure 3. Estimated Overall Survival With 95% CI Band of the Electronic Health Record (EHR) Emulation vs Randomized Clinical Trial (RCT) Stratified by Time

Lines depict the survival curves for time to death obtained from Kaplan-Meier survival analysis for the RCT and mean estimated overall survival probability based on doubly robust estimation; shaded areas indicate 95% CIs. We report the results based on the density ratio estimation in Table 2.
that LAC is a viable alternative to OC, leading to an increased adoption rate of LAC and surgeon training opportunities and proficiency.

The observed improvement in patients’ overall survival in the EHR cohort after 2006 vs the RCT cohort before 2001 and within the EHR cohort for early vs later temporal periods may be partially attributed to the development of cancer treatment over the years. For patients with early-stage colon cancer, studies have shown the associations of survival with surgical technical skills, adjuvant chemotherapy, and mutations. These factors might have suggested the pathway to better survival with better training for surgeons, increased adoption of adjuvant chemotherapy, and expanded options and lines for targeted therapies. Recurrence, a common precursor of mortality, is also reported to decrease.

Our temporal trend–adjusted analysis is applicable to clinical evidence studies of therapies for other diseases. Existing studies in colorectal cancer, diabetes, leukemia, and other cancers using EHR data can span years; however, no adjustment for potential temporal trends was reported. Existence of temporal trends raises doubts regarding the use of an external control group whose time span differs substantially from that of the experimental group. The issue of temporal trends has been raised in previous literature, but the solution until now has been simplistic patient exclusion with potential sampling bias. In the spirit of our approach of stratification by calendar year, the cotrained time-dependent external control can be considered an alternative approach.

The knowledge of findings from the target trial might be associated with the design of the EHR emulation. Franklin et al considered prespecifying the analysis plan to reduce subjective bias, but they only involved structured data. The experience on incorporation of natural language–processing tools and temporal trend adjustment from our study may guide future analysis plans with improved granularity.

**Limitations**

Our study has limitations. First, treatment by indication bias may exist for the analysis during the 2006 to 2009 temporal period. With the extensive adjustment for confounding, the LAC overall survival is still visibly better than OC overall survival within the first 2 years. We hypothesize that only healthier patients underwent LAC from pioneering surgical teams because new procedures tend to be adopted cautiously. Treatment by indication could be caused by a small number of surgical teams capable of performing LAC during the period. The information regarding the indication mechanism might be contained in detailed radiological characteristics but not extractable for our analysis. Caution should be taken for EHR data collected close to the therapy approval date. Second, missing data on radiological tumor characteristics and stage in EHR may lead to inaccurate matching of the RCT eligibility criteria. We developed a rule-based extraction for tumor location, but validation was infeasible because of the lack of criterion standard information. We were unable to replicate the criterion for tumor size owing to the absence of exact information in EHR medical reports. Patients

<table>
<thead>
<tr>
<th>Time</th>
<th>Estimated 5-y survival (95% CI)</th>
<th>OC group</th>
<th>LAC group</th>
<th>Average treatment effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHR emulation cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.847 (0.805 to 0.890)</td>
<td>0.841 (0.793 to 0.888)</td>
<td>-0.006 (−0.070 to 0.057)</td>
<td></td>
</tr>
<tr>
<td>2006-2009</td>
<td>0.767 (0.694 to 0.837)</td>
<td>0.819 (0.705 to 0.922)</td>
<td>0.052 (−0.076 to 0.166)</td>
<td></td>
</tr>
<tr>
<td>2010-2013</td>
<td>0.855 (0.791 to 0.907)</td>
<td>0.834 (0.767 to 0.905)</td>
<td>-0.021 (−0.101 to 0.080)</td>
<td></td>
</tr>
<tr>
<td>2014-2017</td>
<td>0.885 (0.803 to 0.948)</td>
<td>0.857 (0.771 to 0.923)</td>
<td>-0.029 (−0.135 to 0.071)</td>
<td></td>
</tr>
<tr>
<td>Difference between EHR emulation and RCT cohorts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.074 (0.011 to 0.131)</td>
<td>0.061 (-0.001 to 0.122)</td>
<td>-0.013 (−0.077 to 0.093)</td>
<td></td>
</tr>
<tr>
<td>2006-2009</td>
<td>-0.006 (-0.090 to 0.079)</td>
<td>0.039 (-0.085 to 0.152)</td>
<td>0.045 (−0.183 to 0.100)</td>
<td></td>
</tr>
<tr>
<td>2010-2013</td>
<td>0.083 (0.006 to 0.151)</td>
<td>0.055 (-0.025 to 0.140)</td>
<td>-0.028 (−0.092 to 0.121)</td>
<td></td>
</tr>
<tr>
<td>2014-2017</td>
<td>0.112 (0.019 to 0.193)</td>
<td>0.077 (-0.016 to 0.158)</td>
<td>-0.035 (−0.083 to 0.154)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: EHR, electronic health record; LAC, laparoscopy-assisted colectomy; OC, open colectomy; RCT, randomized clinical trial.

* For each analysis, we applied 3 adjustments: outcome regression, propensity scores, and density ratio.

* With the OC group as reference, a positive difference in the 5-year overall survival indicates higher survival probability with OC.

* The difference is calculated between 5-year overall survival estimates of OC, LAC, and mean treatment effect for the EHR vs RCT cohorts.
with missing information on stage were also excluded. Third, physician and surgeon information are not available for the study owing to privacy rules. The proficiency of the surgical team is essential for the quality of the colectomy and may affect the survival through short-term complications and long-term recurrence.39–41 Validation of the analysis on a similar RCT is limited by data availability. These limitations await future study with imaging recognition, special privacy approvals, and advanced methods.

Conclusions

In this comparative effectiveness study, confounding bias due to temporal trends is a major concern for clinical evidence studies spanning several years. Stratification of calendar time and cotraining of models allowed us to account for these temporal trends in models using EHR data. With the proper adjustment, clinical evidence may supplement RCT in assessment of treatment outcomes as well as its evolution over time.

ARTICLE INFORMATION

Accepted for Publication: April 26, 2022.
Published: June 23, 2022. doi:10.1001/jamanetworkopen.2022.18371
Open Access: This is an open access article distributed under the terms of the CC-BY-NC-ND License. © 2022 Hou J et al. JAMA Network Open.

Corresponding Authors: Tianxi Cai, ScD, Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck St, Room 434, Boston, MA 02115 (tcai@hsph.harvard.edu).

Author Affiliations: Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts (Hou, Xiong, Bonzel, Tianxi Cai); Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada (Zhao); Division of Rheumatology, Inflammation, and Immunity, Department of Medicine, Brigham and Women’s Hospital/Harvard Medical School, Boston, Massachusetts (Tianrun Cai, Seyok, Dahal, Liao); Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts (Beaulieu-Jones, Liao, Tianxi Cai); Department of Environmental Health, Harvard T. H. Chan School of Public Health, Boston, Massachusetts (Yuan, Christiani); Merck & Co, Inc, Kenilworth, New Jersey (Fox, Jemielita, Liaw).

Author Contributions: Drs Hou and Tianxi Cai had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Hou, Zhao, Tianrun Cai, Beaulieu-Jones, Fox, Jemielita, Liao, Liaw, Tianxi Cai.

Acquisition, analysis, or interpretation of data: Hou, Zhao, Tianrun Cai, Beaulieu-Jones, Seyok, Dahal, Yuan, Xiong, Bonzel, Fox, Christiani, Liao, Liaw, Tianxi Cai.

Drafting of the manuscript: Hou, Zhao, Tianrun Cai, Beaulieu-Jones, Dahal, Xiong, Bonzel, Jemielita, Liao, Liaw, Tianxi Cai.

Critical revision of the manuscript for important intellectual content: Hou, Tianrun Cai, Beaulieu-Jones, Seyok, Yuan, Fox, Christiani, Liao, Liaw, Tianxi Cai.

Statistical analysis: Hou, Yuan, Jemielita, Tianxi Cai.

Obtained funding: Beaulieu-Jones, Christiani, Liaw, Tianxi Cai.

Administrative, technical, or material support: Hou, Tianrun Cai, Beaulieu-Jones, Xiong, Bonzel, Fox, Christiani, Jemielita, Liao.


Conflict of Interest Disclosures: Dr Hou reported receiving grants from Merck & Co, Inc, during the conduct of the study. Dr Beaulieu-Jones reported receiving grants from Merck SA during the conduct of the study. Dr Fox reported personal fees from and being an employee of Merck Sharp & Dohme Corp, a subsidiary of Merck & Co, Inc, outside the submitted work. Dr Jemielita reported being an employee of Merck & Co, Inc. Dr Liaw reported being an employee of Merck & Co, Inc. No other disclosures were reported.

Funding/Support: The study was supported by Merck & Co, Inc; by grants P30 AR072577 (Dr Liao) and K99NS114850 (Dr Beaulieu-Jones) from the National Institutes of Health (NIH); and grant U01CA209414 from the National Cancer Institute of the NIH (Dr Christiani).
Role of the Funder/Sponsor: Merck & Co, Inc, contributed to the design and conduct of the study; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication but no role in the collection, management, analysis, and interpretation of the data. The NIH played no role in design and conduct of the study; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication; collection, management, analysis, and interpretation of the data.

Additional Information: This study is based on research using information obtained from www.projectdatasphere.org, which is maintained by Project Data Sphere. Neither Project Data Sphere nor the owner(s) of any information from the website have contributed to, approved, or are in any way responsible for the contents of this report. Code for analysis and figure generation is available at https://github.com/hjmarquis/REST_paper1. Summary data that support the findings of this study as well as anonymous data that were used for this study are available on reasonable request to the corresponding author.

REFERENCES


SUPPLEMENT.

eMethods 1. Details of the Data Curation
eMethods 2. Details of the Statistical Analysis
eResults. Temporal Trends in Identified Confounding Factors and Comparison With Benchmark Analyses
eFigure 1. Study Schematics
eFigure 2. Receptor Operating Characteristics (ROC) for MAP Prediction, Number of Colorectal Cancer Diagnosis Code and Number of Mentions of Colorectal Cancer in Medical Notes
eFigure 3. Concordance Between NLP BMI Extraction and Structured BMI Over Patients With Both Extractions
eFigure 4. Summaries on Follow-up Durations for 943 Patients in EHR Emulation Cohort
eFigure 5. Estimated Overall Survival With 95% Confidence Band From Crude Analysis, Benchmark Analysis, Temporal Effect Adjusted Analysis, and RCT
eTable 1. Key Eligibility Criteria of the Target RCT and Their Correspondence in EHR Data
eTable 2. EHR Variables Forming the Filter for Colorectal Cancer Data Mart