Cares: an interactive tool for case-crossover analyses of electronic medical records for patient safety

Alexandre Caron, Emmanuel Chazard, Joris Muller, Renaud Perichon, Laurie Ferret, Vassilis Koutkias, Regis Beuscart, Jean-Baptiste Beuscart, Gregoire Ficheur

Journal of the American Medical Informatics Association, 24(2), 2017, 323–330
doi: 10.1093/jamia/ocw132
Advance Access Publication Date: 27 September 2016
Research and Applications

Research and Applications

IT-CARES: an interactive tool for case-crossover analyses of electronic medical records for patient safety

Alexandre Caron,[1] Emmanuel Chazard,[1] Joris Muller,[1] Renaud Perichon,[1] Laurie Ferret,[2] Vassilis Koutkias,[3] Régis Beuscart,[1] Jean-Baptiste Beuscart,[4] and Grégoire Ficheur[1]

[1]Department of Public Health, Univ. Lille EA 2694, CHU Lille, F-59000 Lille, France, [2]Department of Pharmacy, CHU Lille, F-59000 Lille, France, [3]Institute of Applied Biosciences, Center for Research and Technology Hellas, Thermi, Thessaloniki, Greece and [4]Department of Geriatric Medicine, Univ. Lille EA 2694, CHU Lille, F-59000 Lille, France

Correspondence to Grégoire Ficheur, 2 avenue Oscar Lambret, F-59037 Lille Cedex, France; gregoire.ficheur@univ-lille2.fr; Tel. +33 7 63 17 00 01

Received 4 December 2015; Revised 19 July 2016; Accepted 16 August 2016

ABSTRACT

Background: The significant risk of adverse events following medical procedures supports a clinical epidemiological approach based on the analyses of collections of electronic medical records. Data analytical tools might help clinical epidemiologists develop more appropriate case-crossover designs for monitoring patient safety.
Objective: To develop and assess the methodological quality of an interactive tool for use by clinical epidemiologists to systematically design case-crossover analyses of large electronic medical records databases.
Material and Methods: We developed IT-CARES, an analytical tool implementing case-crossover design, to explore the association between exposures and outcomes. The exposures and outcomes are defined by clinical epidemiologists via lists of codes entered via a user interface screen. We tested IT-CARES on data from the French national inpatient stay database, which documents diagnoses and medical procedures for 170 million inpatient stays between 2007 and 2013. We compared the results of our analysis with reference data from the literature on thromboembolic risk after delivery and bleeding risk after total hip replacement.
Results: IT-CARES provides a user interface with 3 columns: (i) the outcome criteria in the left-hand column, (ii) the exposure criteria in the right-hand column, and (iii) the estimated risk (odds ratios, presented in both graphical and tabular formats) in the middle column. The estimated odds ratios were consistent with the reference literature data.
Discussion: IT-CARES may enhance patient safety by facilitating clinical epidemiological studies of adverse events following medical procedures. The tool's usability must be evaluated and improved in further research.

Key words: Patient safety, data analytics, adverse event, clinical epidemiology, big data, medical informatics

INTRODUCTION

The risks of adverse events following medical procedures can be assessed (at least in part) in randomized controlled trials (RCTs). However, the external validity of RCTs is limited by the strict eligibility criteria and the short follow-up period.[1] Therefore, a robust assessment of patient safety requires large population-based studies.[2] A clinical epidemiological approach based on electronic medical record (EMR) databases would enable a more systematic analysis of adverse events in routine clinical practice. Since many different designs are available, the Observational Medical Outcome Partnership recently carried out an empirically-based comparison of several designs for observational studies.[3] The project concluded

that case-crossover and cohort-crossover designs were the most suitable for pharmacoepidemiological population-based studies.[4]

In the case-crossover design, the case and the control are one and the same person (albeit at different times); this contrasts with the case-control design. The case-crossover design allows the investigator to control for time-constant confounding factors such as gender, age, weight, and lifestyle patterns. Each patient's likelihood of exposure during the period preceding the onset of a given outcome/event (the case period) can be compared with the likelihood of exposure during a similar period at another time (the control period). This design is notably appropriate for transient events or exposures, such as surgical procedures.

Given the large size of today's administrative databases (including EMRs) and the many potential associations to be studied, there is a need to summarize and represent the information more effectively.[5,6] Other epidemiological fields already use web-based strategies and data visualization methods to seek or predict epidemiological events.[7] The visualization of clinical and public health data for complex study designs can be facilitated by the use of analytical tools.[8–10] The integration of visual analytics with advanced statistical methods can help policy makers make more reliable decisions.[11,12]

Clinical epidemiologists are involved in the validation of the data associated with suspected adverse events.[13] For instance, the US Food and Drug Administration reporting program *MedWatch* receives hundreds of thousands of reports each year.[14] These suspected associations between an exposure (e.g., a medical procedure) and an adverse event should be explored in clinical epidemiological studies. However, this task requires advanced statistical knowledge and skill in effectively locating and exploiting information within large databases.

The specific problem of venous thromboembolism (VTE) and the iatrogenic bleeding risk associated with antithrombotic drugs has been addressed in a number of clinical epidemiological studies.[15–17] Indeed, almost half of all cases of VTE can be attributed to current or recent hospitalization,[18] and risk factors for VTE are very common among hospitalized patients.[19] Around the world, the current guidelines recommend active prevention strategies.[20–22] The RCTs conducted to date have failed to determine the long-term risk of VTE or bleeding,[23] and the generalizability of these RCTs' findings has been limited due to low compliance with care guidelines (poor adherence to dosing, insufficient treatment durations, the prescription of inappropriate antithrombotic, etc.).[24–27]

The present study had 2 objectives: (i) to develop an interactive tool for use by clinical epidemiologists to systematically design case-crossover analyses of large EMR databases and (ii) to assess the tool's applicability and methodological quality by using it to estimate thromboembolic and bleeding risks after medical procedures recorded in a very large nationwide administrative inpatient database.

MATERIAL AND METHODS

We developed "IT-CARES," an interactive tool for the case-crossover analyses of EMRs. We shall successively present the case-crossover design embedded in the tool, the implementation of IT-CARES (including the input dataset and the user interface settings), and the details of our test case (based on the French national inpatient stay database).

A case-crossover design in IT-CARES

A case was defined as a patient's first experience of the primary outcome (e.g., VTE or bleeding). After including the case, we searched the EMR database for exposure in 2 periods: (i) immediately before the primary outcome and (ii) during a control period 1 year earlier (Figure 1). The case period and thus the control period were split into several intervals. A paired-matched interval approach was used, as described by Mittleman et al.[28] Conditional logistic regression was used to compare the likelihood of exposure during each interval of the case period with that of the control period. An odds ratio (OR) and its 95% confidence interval (CI) was computed for each interval. This OR reflects the risk of onset of the primary outcome compared with the baseline risk. We automated the case-crossover analyses by adopting some modeling assumptions and incorporating user-defined parameters into IT-CARES.

Modeling assumptions

The index date (the day of onset of the primary outcome) was assumed to be the first day of the inpatient stay episode. For each case, an observation window of 2 years (from the index date) was defined retrospectively. Once the patient's index date had been determined, the case and control periods were screened for exposure. The "time to case" was defined as the difference between the index date and the exposure date. If the exposure and the primary outcome occurred during the same episode, the time to event was set to 0.

Parameters defined by the user when performing the case-crossover analysis

(a) Criteria for case inclusion: the clinical epidemiologist chose 1 or more diagnoses to define the case's primary outcome and delimit the study period. The index date had to be inside the study period (apart from the first 2 years). The age had to be within a defined range.

(b) Criteria for case exclusion: patients were excluded by the clinical epidemiologist if they had a medical history defined by a list of 1 or more diagnoses. Exclusion criteria were applied to the case stay itself, if required.

(c) Criteria for exposure: the clinical epidemiologist defined the exposure as a combination of diagnoses and medical procedures. The database was searched for this combination during the case period and control period. The clinical epidemiologist also defined the maximum possible length for an episode with exposure. The case stay was excluded from the exposure screening period, if required. Lastly, the epidemiologist could exclude a patient presenting several episodes of the exposure.

Implementation of IT-CARES

Input dataset

We set up a simple data model consistent with the case-crossover design; the input dataset had to comply with a denormalized format containing at least the following 8 columns: (1) patient ID, (2) episode ID, (3) diagnoses, (4) procedures, (5) age, (6) admission day, (7) year of the episode, and (8) length of stay (for details, see Supplementary Appendices). In order to be analyzed with IT-CARES, data from our test case had to be formatted as described; these data are presented below in the *Data sources* section.

User interface settings

The user interface was developed to enable the user (e.g., the clinical epidemiologist) to set the case-crossover parameters presented in Table 1. He/she selected the input database corresponding to the risk to be studied (e.g., VTE or bleeding), set the values of the study

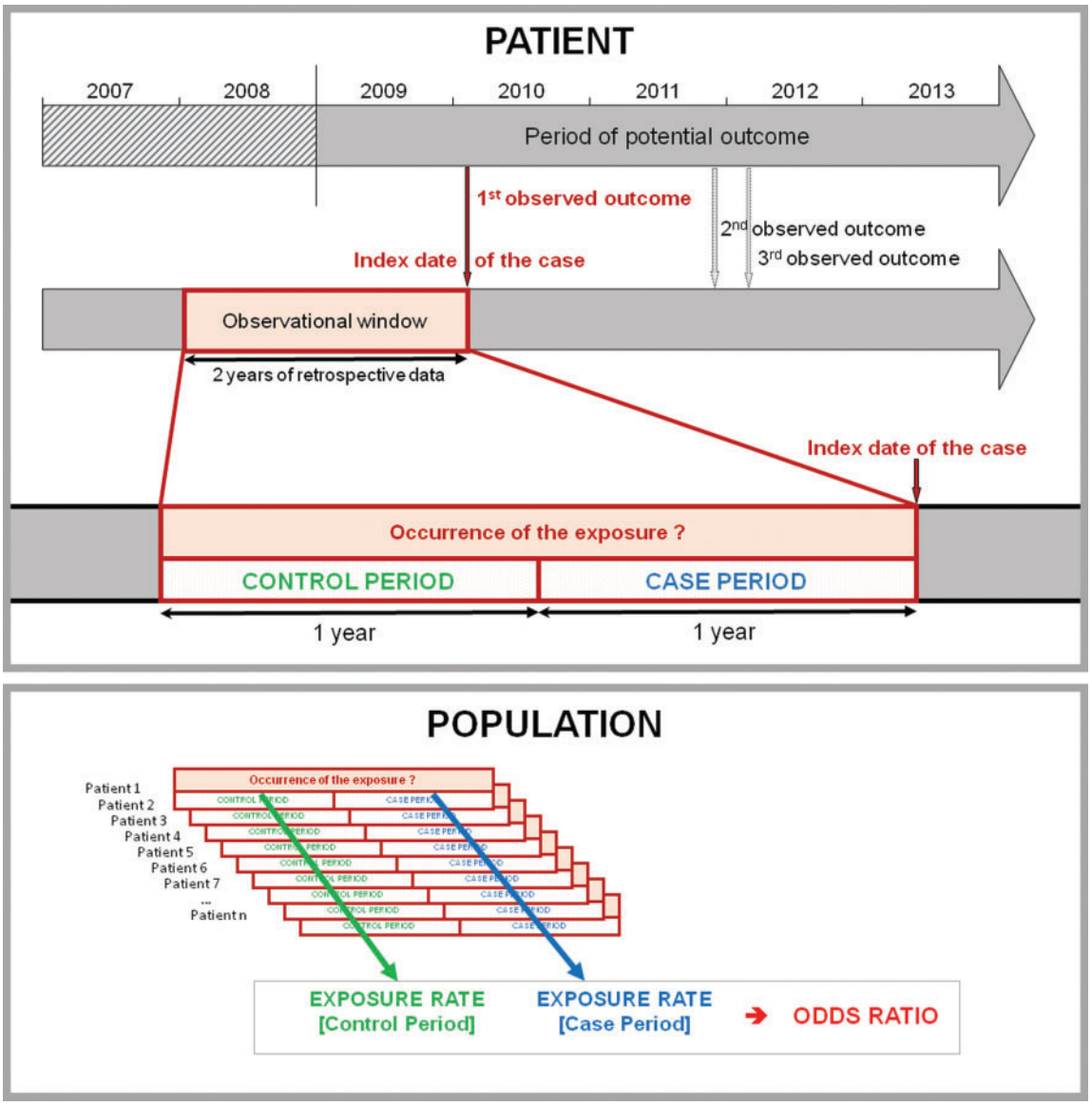

Figure 1. The case-crossover design. Patient: (1) the primary outcome is sought between 2009 and 2013 (the hatched time line, providing 2 years of retrospective data). (2) The first occurrence defines a retrospective observational window of 2 years. (3) The exposure is sought during case and control periods. Population (all cases): rates of exposure during case and control periods are compared using conditional logistic regression, and the odds ratio is computed.

parameters, and queried the database. IT-CARES automatically estimated the risk and generated output in both graphical and tabular formats.

Technical specifications

The IT-CARES software and a set of simulated data are made available via an R package to allow any clinical epidemiologist to analyze his/her own collection of EMRs using a systematic case-crossover design. IT-CARES was implemented using *R 3.2.0*[29] and *RShiny*, a framework for creating interactive web applications based on *R*.[30] Support for processing large databases (i.e., with tens of millions of rows) was optimized with the *data.table*[31] and *dplyr*[32] packages. Final rendering was performed with *ggplot2*,[33] *xtable*,[34] and *markdown*.[35] Statistical analysis used the *survival* package.[36] All analyses were performed on a server with an Intel® Xeon® Processor E5-2620 v2 @ 2.10 GHz chip and 32 Go of DDR3 1600 MHz memory.

Given the large number of records typically used to implement case-crossover studies, computational performance is a challenging issue. Case selection is a time-consuming process because the whole database is queried. Since case/exposure selection is a serial process, any change in the case selection criteria requires the exposure selection to be recomputed. However, the converse is not true, meaning that outputs can be updated faster when exposure selection parameters alone are modified. Computation time is presented with the test case in the Results section.

The test case: using IT-CARES to analyze 170 million inpatient stays

Risks of thromboembolism and bleeding after medical procedures

In the context of our test case, we used IT-CARES to estimate the population-based risk of an acute thromboembolic or bleeding event (the primary outcome) following exposure to a medical procedure.

Table 1. User-defined parameters in IT-CARES for case-crossover analyses

Parameter	Value/value range	Definition and use
Primary outcome	List of diagnostic codes	Select the cases: the earliest record containing 1 or more of these diagnoses
Study period	Start year and end year	Specify the study period
Age	Minimum and maximum age	Check whether the patient's age on the index date is within the range
Exclusion criteria	List of diagnostic codes	Exclude a patient if any of his/her records contain 1 or more of these diagnoses
Application of the exclusion criteria to the case stay	TRUE/FALSE	Apply the exclusion criteria to the case stay, or not
Exposure		Select the exposure: all records containing 1 or more of these codes during the case or control periods
Diagnoses and/or	List of diagnostic codes	
Procedures	List of procedure codes	
Screening case stay	TRUE/FALSE	Screen the case stay for exposure
Only 1 exposure allowed	TRUE/FALSE	Exclude multiple exposures
Maximum length of stay	Length of stay	Limit the length of the episode with exposure
Case/control periods	Length of an interval (in days) Number of intervals	The length of the case and control periods are defined as the length of an interval multiplied by the number of intervals

We used a washout period of 1 year (between the end of the control period and the onset of the primary outcome) to ensure that the risk returned to baseline. Medical procedures are influenced by seasonal trends; thus, a 1-year period took account of the fact that time of year might be a confounding factor. After setting appropriate parameters for each study, the risk was automatically estimated using IT-CARES.

Data sources

The French national inpatient stay database contains an exhaustive structured description of all inpatient stays in French public- and private-sector hospitals. It was first designed for health insurance payment purposes. We reused the "acute hospital admissions" part of the database, which contains 171 556 421 inpatient stays for the 7-year period from January 1, 2007, to December 31, 2013. Each record contained data on the diagnoses (according to the International Statistical Classification of Diseases and Related Health Problems, 10th Revision, ICD-10), medical procedures (according to the French *Classification Commune des Actes Médicaux* classification), age, year of the discharge, length of stay, time between 2 admissions, and a unique patient identifier. Although the database is pseudo-anonymized, the unique identifier allowed us to track individual patients in the database and compile each inpatient stay within the study period. We preprocessed the database by selecting a subset that contained all stays for patients with at least 1 diagnosis of thromboembolism or bleeding (for the codes, please refer to the Supplementary Appendices). The study was approved by the French data protection authority (Commission Nationale de l'Informatique et des Libertés (CNIL) authorization number: 1754053).

Comparison of IT-CARES outputs with relevant published studies

The results generated by IT-CARES were compared with relevant published studies in a 3-step process. We first replicated Kamel et al.'s findings on the risk of thrombotic events after delivery.[17] In Kamel et al.'s study, the risk was computed using a retrospective crossover-cohort analysis on claims data for discharges from acute care hospitals and emergency departments in California. A composite primary outcome of ischemic stroke, acute myocardial infarction, or VTE was used, although the researchers also provided ORs for VTE alone. We thus compared the ORs computed with IT-CARES with the ORs for the composite outcome and for VTE computed by Kamel et al. Cases of thromboembolism were screened for using pulmonary embolism (PE) as the primary outcome. We included patients aged between 15 and 45 years. Patients with a history of VTE were excluded. We screened the cases for codes for delivery as the exposure. We did not screen for exposure in the case stay and allowed only 1 exposure. We computed seven 42-day intervals.

Second, we assessed the bleeding risk after total hip replacement (THR) and compared it with the value from Lalmohamed et al.'s[15] study of Danish national registry data. The latter database included (but was not limited to) information on hospital stays, outpatient visits, drugs, and death. Using a Cox proportional hazards model, Lalmohamed et al. computed the hazard ratio (HR) for gastrointestinal bleeding in THR patients, relative to age- and gender-matched controls. The researchers used time interaction terms to estimate HRs for different periods. We identified bleeding using the same list of ICD-10 codes. Cases of bleeding were screened for using serious intestinal bleeding as the primary outcome. We included patients aged 18 or over. Patients with a history of serious intestinal or intracranial bleeding were excluded. We screened the cases for procedure codes for THR (exposure). We did not screen the case stay and allowed only 1 exposure. We computed six 42-day intervals.

Lastly, since carpal tunnel surgery reportedly does not increase the risk of VTE, this procedure was chosen as a negative control.

RESULTS

In the following sections, we present the IT-CARES user interface, its availability (as an open-source tool), and the results of the test case.

Presentation of the user interface

The deployment of IT-CARES as a web application is depicted in Figure 2. The user interface is divided into 3 columns. The left and right columns are dedicated to user inputs, whereas the middle column is dedicated to displaying the IT-CARES output. More specifically, the criteria for case selection and the criteria for exposure selection are provided in the left- and right-hand columns, respectively. The middle column is divided into 3 panels: the update button for generating new estimates (top), the graphical output (middle), and the tabular output (bottom). In our test case, the case selection process took an average of 22 s (range: 20–27 s) and exposure selection took an average of 4 s (range: 2–7 s).

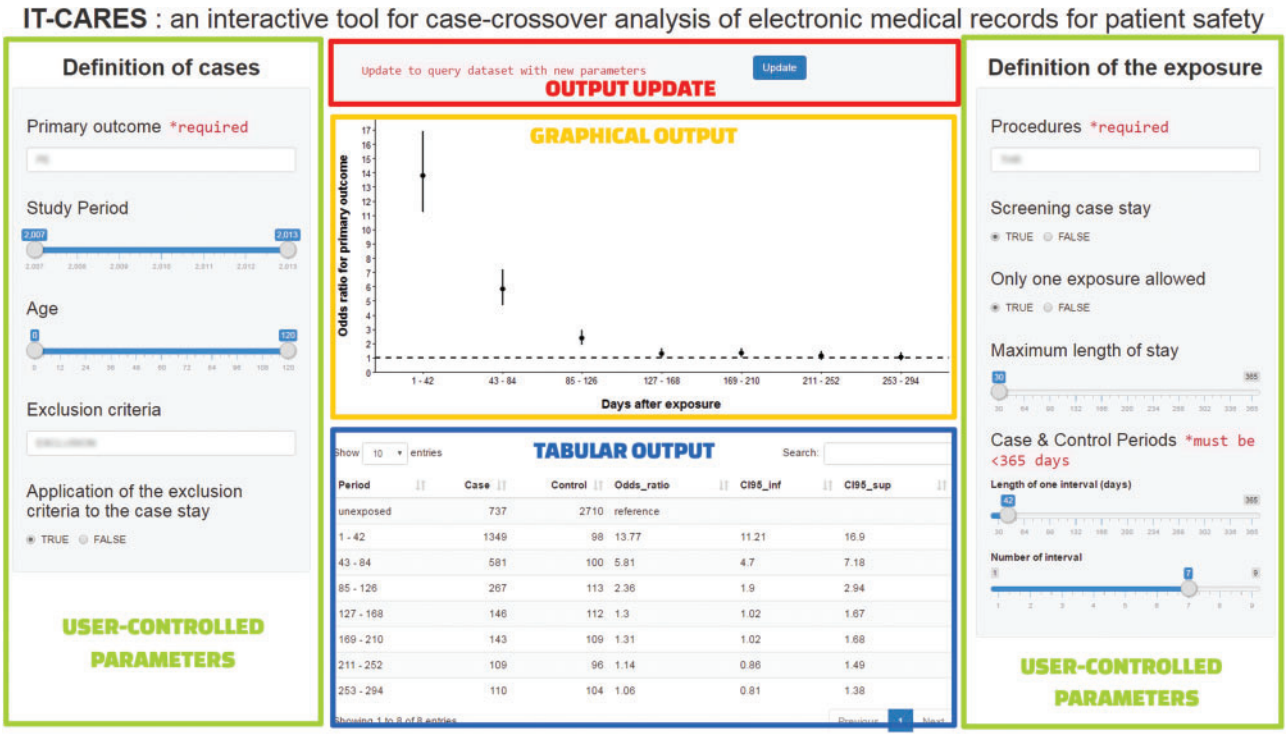

Figure 2. Web application deployment of IT-CARES; user-controlled parameters (green fields), and graphical and table outputs (orange and blue fields).

Availability of IT-CARES

IT-CARES will be available from the Comprehensive R Archive Network (*https://cran.r-project.org/*) as an R package. The source code is freely available on GitHub (https://github.com/jomuller/ITCARES). IT-CARES was also deployed as an interactive web application through RShiny. Thus, IT-CARES outputs can be generated through the web application or by calling its methods in the R console.

The test case: automated risk estimations and comparisons with relevant published studies

We first configured IT-CARES to estimate the risk of a thromboembolic event after delivery. We identified 231 264 first cases of PE. As shown in Figure 3, we observed 410 exposures during the first interval after delivery (42 days) in the case period and 36 exposures within the same interval in the control period. The risk of thromboembolism was significantly elevated during this interval, with an OR [95% CI] of 11.39 [8.10–16.01]. During the same period, Kamel et al. found a thromboembolic risk (expressed as an OR [95% CI]) of 10.8 [7.8–15.1]. The risk was also elevated during the second interval of 42 days, and both studies showed that the thromboembolic risk was not elevated after 12 weeks (Table 2).

In a second step, we configured IT-CARES to estimate the bleeding risk after THR. We identified 515 580 first cases of serious intestinal bleeding. As shown in Figure 4, we observed (i) 576 exposures within the first interval of 42 days during the case period and (ii) 109 exposures within the same time interval during the control period. The associated risk of serious bleeding was significantly elevated during this interval, with an OR [95% CI] = 5.3 [4.3–6.5]. Lalmohamed split the first interval into days 1–14 and days 15–42 (Table 2). The HRs [95% CI] for the bleeding risk were, respectively, 6.0 [4.1–8.9] and 4.3 [3.3–5.7]. The risk was also elevated during the second 42-day interval, and both studies showed that the risk of serious bleeding persisted for 12 weeks.

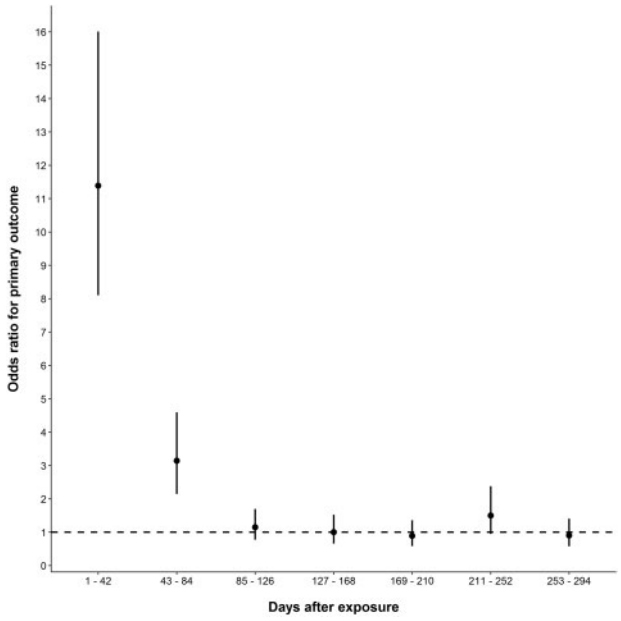

Figure 3. IT-CARES's graphical presentation of the risk of venous thromboembolism in successive 42-day intervals after delivery.

A complementary analysis was performed as a negative control. As expected, the risk of a thromboembolic event after carpal tunnel surgery was not significantly elevated during the 8-month study period (see the Figure in the Supplementary Appendices).

Table 2. Comparison of our estimations with the findings of relevant published studies

Interval	Risk of thromboembolic event after delivery (Kamel et al.)			Risk of serious bleeding after total hip replacement (Lalmohamed et al.)	
	IT-CARES (PE[a])	Kamel (overall[b])	Kamel (VTE[c])	IT-CARES	Lalmohamed
Interval 1 (days 1–42)	**11.4 [8.1–16.0]**	10.8 [7.8–15.1]	12.1 [7.9–18.6]	**5.3 [4.3–6.5]**	6.0 [4.1–8.9] (days 1–14) 4.3 [3.3–5.7] (days 15–42)
Interval 2 (days 43–84)	**3.1 [2.1–4.6]**	2.2 [1.5–3.1]	2.2 [1.4–3.3]	**1.9 [1.5–2.3]**	2.4 [1.8–3.2]
Interval 3 (days 85–126)	**1.2 [0.8–1.7]**	1.4 [0.9–2.1]	1.6 [1.0–2.5]	**1.2 [0.9–1.6]**	1.0 [0.8–1.3] (days 85–182)

The results are quoted as the OR [95% CI] with the exception of Lalmohamed et al.'s study, for which the HR [95% CI] is given.
Emphasis of our results (IT-CARES) versus reference studies are indicated in bold.
[a]Primary outcome = pulmonary embolism alone.
[b]Primary outcome = stroke + myocardial infarction + VTE.
[c]Primary outcome = VTE.

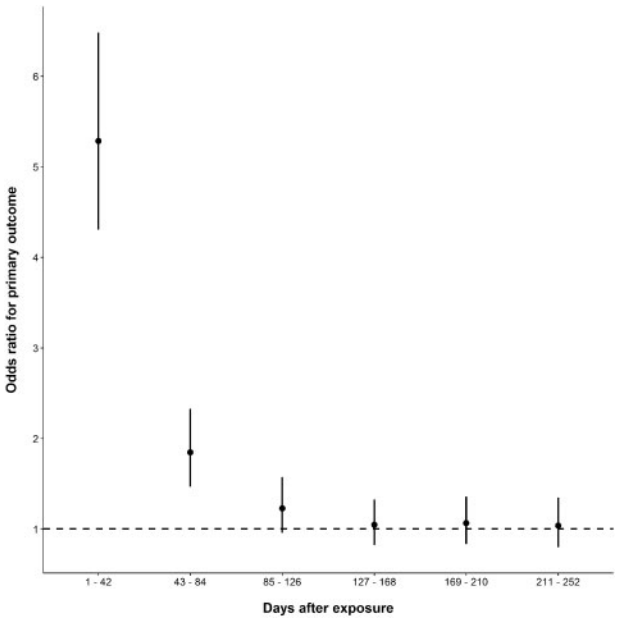

Figure 4. IT-CARES's graphical presentation of the risk of serious bleeding in successive 42-day intervals after total hip replacement.

DISCUSSION

Main findings

IT-CARES was developed as an interactive analytical tool that enables clinical epidemiologists to design and perform a case-crossover analysis exploiting a large collection of EMRs. We demonstrated the tool's capabilities and accuracy in 2 test assessments of the population-based thromboembolic risk after delivery and the bleeding risk after THR, respectively. IT-CARES makes it possible for clinical epidemiologists to design and rapidly execute a complex case-crossover analysis in a very large database. Clinical epidemiologists are likely to want to explore a wide range of medical procedures and potential outcomes. These associations can be addressed by redefining the primary outcome and appropriate design settings in IT-CARES. Moreover, IT-CARES makes it easy to assess the trade-off between different risks of adverse events (such as bleeding and thrombosis in the test case presented in our study), which can be studied over different periods of time after medical procedures. Given the lack of other similar analytical tools, we compared IT-CARES's results with relevant published studies. Our results for the thromboembolic and bleeding risks are consistent with the findings reported by Kamel et al.[17] and Lalmohamed et al.[15] in terms of both the effect size and the persistence of risk over time. We also performed a negative control (carpal tunnel surgery); as expected, we did not observe a significant elevation of the thromboembolic risk after this day-case surgery.

Strengths and limitations of IT-CARES

IT-CARES relies on the exploitation of EMR data. The use of EMR databases has already revealed opportunities for improving patient safety and the quality of care.[37,38] However, many concerns about the reliability of the information in EMRs have been expressed.[39] These potential limitations are closely related to the design of a system that generally focuses on a specific aspect such as cost, efficiency, quality, or patient safety.[40] Many different types of EMR databases are found across the world, which leads to poor interoperability and contrasting data models.[41] This is one of the reasons we chose to publish the code for IT-CARES and selected a simple, denormalized format for the data model. Reusing administrative data is advantageous, insofar as this approach provides a very large sample size and is relatively inexpensive (with a moderate marginal cost).[42] Furthermore, subgroups of patients with particular characteristics can be targeted in order to evaluate a specific risk and its persistence over time. For a given medical risk, IT-CARES could be effectively used to estimate changes over time or to compare these risks from one hospital to another. This would be of great interest to health authorities or hospitals *per se*. Furthermore, the integration of an administrative database with drug prescription data would enable IT-CARES to assess drug safety–related risk.

Given the inherent complexity of implementing case-crossover study designs, IT-CARES should not be considered the sole means of conducting these studies. In fact, it is a tool that can help an investigator conduct case-crossover analyses in a more systematic way. Thus, clinical epidemiologists will need to be trained in the routine use of IT-CARES.[5] This also implies that IT-CARES's usability needs further testing. One of IT-CARES's key features for the clinical epidemiologist is the ability to adjust the observation window and case/control periods. Nonetheless, users have to be cautious when comparing the case period with the usual length of stay associated with the medical procedure. A case period shorter than the length of stay would lead to overestimation of the first OR by erroneously including late cases. This problem could be solved by replacing the admission day with the exact start day for the primary outcome onset—if the latter is available in the EMR dataset. The case-crossover design also requires the onset of the case to be acute;

the time between exposure and onset of the primary outcome must therefore be accurately computed.

Lessons learned from our test case

Our test case emphasized a number of issues that have to be taken into account by the clinical epidemiologist.[43] We did not know the precise date of event occurrence within a stay. Consequently, if the exposure and the primary outcome are present in the same record, we did not know which occurred first. However, we assumed that most medical procedures would be postponed if the primary outcome occurred first—making it highly probable that exposure preceded the outcome in the cases we analyzed. This assumption is not valid for delivery, which by definition is rarely deferrable; hence, we excluded the case stay when screening for exposure in the test case. Another point mentioned above is the reliability of coding medical data. Our database allowed us to use only PE and a limited number of bleeding codes, since this event has a sudden onset and is almost always serious enough to require hospitalization.[44,45] Lastly, our database did not record deaths outside the hospital, which might have led to the underestimation of associations.

Moreover, a difference in the way the primary outcome is measured in the case period vs the control period would introduce a classification bias. For instance, diagnosis of PE can be overestimated after surgical exposure because the clinicians are aware of the high risk and are more likely to look for signs of this condition. Likewise, better outpatient follow-up might lead to overestimation of the association.

CONCLUSION

IT-CARES was developed as an interactive, freely-available, open-source tool enabling the clinical epidemiologist to implement the case-crossover design systematically in EMR databases. This tool may enhance patient safety by facilitating adverse event assessment studies following medical procedures. Although IT-CARES provided reliable results in a test case, further research must be carried out in order to evaluate it in additional patient safety studies and elaborate on its usability for advancing the end-user experience.

CONTRIBUTORS

Conception and design of the study: GF, AC, and EC. Assisted with the implementation of statistical analysis: RB, EC, and GF. Assisted with the analysis of patient safety: EC, LF, and VK. Development, testing and the web version of IT-CARES: AC, RP, JM, and GF. Package development: AC and JM. Led the writing of the paper: AC. Contributed to revision of the paper: GF, JB, EC, VK, JM, RP, LF, and RB. All authors approved the manuscript's results and conclusions.

ACKNOWLEDGMENTS

None.

FUNDING

The present study was funded solely by the University of Lille and Lille University Medical Center (Lille, France).

COMPETING INTERESTS

None.

ETHICS APPROVAL

This study was authorized by the French data protection authority (*Commission Nationale de l'Informatique et des Libertés*, CNIL), reference 1 754 053.

SUPPLEMENTARY MATERIAL

Supplementary material are available online at *Journal of the American Medical Informatics Association*.

REFERENCES

1. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" *Lancet* 2005;365:82–93.
2. Strom BL, Kimmel SE, Hennessy S. *Pharmacoepidemiology*. John Wiley & Sons; 2011.
3. Evans SJW. Moving along the yellow brick (card) road? *Drug Saf* 2013;36:3–4.
4. Ryan PB, Stang PE, Overhage JM, *et al*. A comparison of the empirical performance of methods for a risk identification system. *Drug Saf* 2013;36 (Suppl 1):S143–S158.
5. Thomas J, Kielman J. Challenges for visual analytics. *Inf Vis* 2009;8:309–14.
6. Margolis R, Derr L, Dunn M, *et al*. The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. *J Am Med Inform Assoc* 2014;21:957–8.
7. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009;11:e11.
8. Shortliffe EH, Cimino JJ, editors. *Biomedical Informatics*. London: Springer; 2014. Available at: http://link.springer.com/10.1007/978-1-4471-4474-8. Accessed 5 August 2015.
9. HOMER for population-level estimation | OHDSI. Available at: http://www.ohdsi.org/analytic-tools/homer-for-population-level-estimation/. Accessed 8 October 2015.
10. Caban JJ, Gotz D. Visual analytics in healthcare – opportunities and research challenges. *J Am Med Inform Assoc* 2015;22:260–2.
11. Shneiderman B, Plaisant C, Hesse BW. Improving healthcare with interactive visualization. *Computer* 2013;46:58–66.
12. EventFlow. Available at: http://www.cs.umd.edu/hcil/eventflow/. Accessed 8 August 2015.
13. World Health Organization. *Patient Safety Curriculum Guide: Multi-professional Edition*. Geneva: World Health Organization; 2011.
14. Kessler DA, Natanblut S, Kennedy D, *et al*. Introducing medwatch: a new approach to reporting medication and device adverse effects and product problems. *JAMA* 1993;269:2765–8.
15. Lalmohamed A, Vestergaard P, Javaid MK, *et al*. Risk of gastrointestinal bleeding in patients undergoing total hip or knee replacement compared with matched controls: a nationwide cohort study. *Am J Gastroenterol* 2013;108:1277–85.
16. Sweetland S, Green J, Liu B, *et al*. Duration and magnitude of the postoperative risk of venous thromboembolism in middle aged women: prospective cohort study. *BMJ* 2009;339:b4583.
17. Kamel H, Navi BB, Sriram N, *et al*. Risk of a thrombotic event after the 6-week postpartum period. *N Engl J Med* 2014;370:1307–15.
18. Heit JA, O'Fallon W, Petterson TM, *et al*. Relative impact of risk factors for deep vein thrombosis and pulmonary embolism: a population-based study. *Arch Intern Med* 2002;162:1245–8.
19. Anderson FA. Risk factors for venous thromboembolism. *Circulation* 2003;107:9I–16.
20. Guyatt GH, Akl EA, Crowther M, *et al*. Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of chest physicians evidence-based clinical practice guidelines. *Chest* 2012;141:7S–47S.
21. Venous thromboembolism: reducing the risk | Guidance and guidelines | NICE. Available at: http://www.nice.org.uk/guidance/cg92. Accessed 16 January 2015.

22. Samama C-M, Albaladejo P, Laversin S, *et al*. Prévention de la maladie thromboembolique veineuse périopératoire et obstétricale. *Ann Fr Anesth Réanimation* 2005;24:853–61.
23. Kakkar AK, Brenner B, Dahl OE, *et al*. Extended duration rivaroxaban versus short-term enoxaparin for the prevention of venous thromboembolism after total hip arthroplasty: a double-blind, randomised controlled trial. *The Lancet* 2008;372:31–9.
24. Cohen AT, Tapson VF, Bergmann J-F, *et al*. Venous thromboembolism risk and prophylaxis in the acute hospital care setting (ENDORSE study): a multinational cross-sectional study. *Lancet Lond Engl* 2008;371:387–94.
25. Grant PJ, Flanders SA. Venous thromboembolism prophylaxis: a path toward more appropriate use. *BMJ Qual Saf* 2015;24:605–7.
26. Wilke T, Müller S. Nonadherence in outpatient thromboprophylaxis after major orthopedic surgery: a systematic review. *Expert Rev Pharmacoecon Outcomes Res* 2010;10:691–700.
27. Warwick D, Friedman RJ, Agnelli G, *et al*. Insufficient duration of venous thromboembolism prophylaxis after total hip or knee replacement when compared with the time course of thromboembolic events findings from the Global Orthopaedic Registry. *J Bone Joint Surg Br* 2007;89-B:799–807.
28. Mittleman MA, Maclure M, Robins JM. Control sampling strategies for case-crossover studies: an assessment of relative efficiency. *Am J Epidemiol* 1995;142:91–8.
29. *R Core Team (*2015*). R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/ Accessed 17 June 2015.
30. Chang W, Cheng J, Allaire JJ, *et al*. *Shiny: Web Application Framework for R*. 2015. Available at: http://cran.r-project.org/web/packages/shiny/index.html. Accessed 15 June 2015.
31. Dowle M, Short T, Lianoglou S, *et al*. *Data.table: Extension of Data.frame*. 2014. Available at: http://cran.r-project.org/web/packages/data.table/index.html. Accessed 17 June 2015.
32. Wickham H, Francois R, RStudio. *Dplyr: A Grammar of Data Manipulation*. 2015. Available at: http://cran.r-project.org/web/packages/dplyr/index.html. Accessed 17 June 2015.
33. Wickham H, Chang W. *ggplot2: An Implementation of the Grammar of Graphics*. 2015. Available at: http://cran.r-project.org/web/packages/ggplot2/index.html. Accessed 17 June 2015.
34. Dahl DB. *xtable: Export tables to LaTeX or HTML*. 2014. Available at: http://cran.r-project.org/web/packages/xtable/index.html. Accessed 17 June 2015.
35. Allaire JJ, Horner J, Marti V, *et al*. *Markdown: "Markdown" Rendering for R*. 2015. Available at: http://cran.r-project.org/web/packages/markdown/index.html. Accessed 17 June 2015.
36. Therneau TM. *Survival Analysis*. 2 July 2015. Available at: http://CRAN.R-project.org/package=survival. Accessed 18 September 2015.
37. McGinn TG, McCullagh L, Kannry J, *et al*. Efficacy of an evidence-based clinical decision support in primary care practices: a randomized clinical trial. *JAMA Intern Med* 2013;173:1584–91.
38. Jones SS, Rudin RS, Perry T, *et al*. Health information technology: an updated systematic review with a focus on meaningful use. *Ann Intern Med* 2014;160:48–54.
39. Oxentenko AS, West CP, Popkave C, *et al*. Time spent on clinical documentation: a survey of internal medicine residents and program directors. *Arch Intern Med* 2010;170:377–80.
40. Payne TH, Corley S, Cullen TA, *et al*. Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. *J Am Med Inform Assoc* 2015;22:1102–10.
41. Vest JR, Zhao H, Jasperson J, *et al*. Factors motivating and affecting health information exchange usage. *J Am Med Inform Assoc* 2011;18:143–9.
42. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care* 2003;12 (Suppl 2):ii58–63.
43. Reich CG, Ryan PB, Schuemie MJ. Alternative outcome definitions and their effect on the performance of methods for observational outcome studies. *Drug Saf* 2013;36:181–93.
44. Hippisley-Cox J, Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleed with anticoagulants: cohort study to derive and validate the QBleed scores. *BMJ* 2014;349:g4606.
45. Casez P, Labarère J, Sevestre M-A, *et al*. ICD-10 hospital discharge diagnosis codes were sensitive for identifying pulmonary embolism but not deep vein thrombosis. *J Clin Epidemiol* 2010;63:790–7.