Bringing cohort studies to the bedside: framework for a ‘green button’ to support clinical decision-making

When providing care, clinicians are expected to take note of clinical practice guidelines, which may not apply to individual patients with comorbidities, as they are typically excluded from clinical trials. Guidelines also tend not to provide relevant evidence on risks, secondary effects and long-term outcomes. Querying the electronic health records of similar patients may for many provide an alternate source of evidence to inform decision-making. It is important to develop methods to support these personalized observational studies at the point-of-care, to understand when these methods may provide valid results, and to validate and integrate these findings with those from clinical trials.

Keywords: cohort studies • comparative effectiveness • electronic health records • personalized prognosis

The demand for information on treatment options and what to expect from them in terms of their effectiveness and safety has grown tremendously, driven by an increase in patients with complex comorbidities as well as in the choice and sophistication of treatments. When providing care, clinicians have limited time to diagnose the presenting problem and provide the patient with a full range of treatment options and associated prognoses. Clinicians are expected to take note of clinical practice guidelines, which offer recommendations based on the available evidence. However, guidelines may not apply to individual patients with comorbidities, as they are typically excluded from clinical trials. Guidelines also tend not to provide relevant evidence on risks, secondary effects and long-term outcomes. Additionally, patients have their own preferences, are unfamiliar with guidelines and are uncertain about the way probabilities of benefit and harm for interventions apply to them.

Querying the electronic health records (EHRs) of similar past patients may for many provide an alternate source of evidence to inform decision-making. Past records are a source of information on the way different treatment choices led to different outcomes and can help in the tailoring of medical treatment to the individual characteristics of each patient. These virtual cohorts of past patients are more likely to represent a realistic population with similar comorbidities than those assembled for clinical trials. As long as the limitations of observational analysis are acknowledged, such cohorts can serve as an important adjunct tool in clinical decision-making.

The idea of systematically searching for ‘patients like mine’ in datasets of patients from clinical practices was first reported in the 1970s. A 1972 paper by clinicians from Veterans Administration hospitals reported the use of a manually built electronic ‘library of clinical experience’ of 678 lung cancer patients for personalized prognosis information for new patients [5]. “Without delegating any of the prerogatives of clinical judgment to the computer,” they state, “a clinician can obtain a quantified account, in as much or as little detail as he wishes, of the experience with previous patients.” Another study in
1975 describes the creation of one of the first electronic patient registries, a dataset of more than 3000 patients with ischemic heart disease [6]. The dataset was used to produce prognostic reports tailored to specific patients, and researchers claimed that: “its use to answer general questions is not sufficient to justify its existence. It must be used to answer specific questions about specific patients.” Since then, important advances from computer science, statistics and the big data revolution have made it possible to implement such decision-support tools at the point-of-care. The first reported use of querying medical records of past patients in near real-time to aid in treatment decision took place in 2011, when clinicians at Stanford University searched for patients similar to a 13 year old girl with lupus nephritis to decide on anticoagulation therapy [7].

We envision future EHR systems will bring querying, visualization and decision support functions together to allow individualized virtual cohorts of similar patients to be assembled and used in real-time to support treatment selection and planning (see Table 1). The policy and ethical implications of such a system were recently considered by a group of authors who branded this EHR capability the ‘green button’ [8].

Operational issues associated with this type of decision support systems have recently begun to be discussed in the literature [9,10]. Here, we outline an approach for the dynamic creation of virtual cohorts in EHR data that combines EHR querying systems with statistical methods from observational studies (see Figure 1). This approach includes: EHR-based phenotyping, to characterize patients; measuring inter-patient similarity, to form cohorts of similar patients; optimal cohort selection, to tailor clinical decisions to individual patients; cohort visualization to facilitate face validity and refinement by the decision maker; automated confounder control, to minimize bias for validity and integrating results with clinical guidelines and existing evidence from clinical trials for validity as well as to fill inferential gaps.

**EHR-based phenotyping**

In cohort studies, correct identification of patients, their treatments and outcomes are important for the estimation of treatment effects. EHR data can be inaccurate or incomplete and can contain codes that classify patients for purposes other than clinical care, distorting meaningful interpretation [11]. Important information is often contained in free-text notes requiring text interpretation [12] and missing data are common and require appropriate inference methods [13].

EHR-based phenotyping algorithms are being developed that can transform high-dimensional, noisy, structured and unstructured data from laboratories, medications, diagnosis codes, procedure codes and clinical notes into meaningful clinical concepts using knowledge representation, temporal abstraction algorithms, natural language processing and machine learning techniques. Examples of these algorithms include identification of patients with diabetes, cancer, heart failure, rheumatoid arthritis, pneumonia, asthma, hypertension, venous thromboembolism and drug adverse events among others. A review of these applications and methodologies can be found in [14]. While there is still

<table>
<thead>
<tr>
<th>Ideal properties of future EHR systems</th>
<th>Current EHR tools</th>
<th>Observational studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Represent real world populations and routine care</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Low marginal cost for additional analyses</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Can look at long term effects and rare conditions</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Capability of real-time, interactive querying</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Appropriate inclusion of sequential nature of healthcare events</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Includes visualization techniques to help with the study design</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Appropriately address measured confounders</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Take into account statistics of multiple testing</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Appropriately address unmeasured or unknown confounders</td>
<td>No</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Data are high quality and fit for purpose</td>
<td>No</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Allows comparison with existing evidence</td>
<td>No</td>
<td>Sometimes</td>
</tr>
</tbody>
</table>

EHR: Electronic health record.
work to be done, particularly around standardization and the use of unsupervised methods to reduce manual intervention, initial results are promising [15].

**Quantifying inter-patient similarity**
Patient attributes (or phenotypes) generated by EHR-based phenotyping algorithms (when required) can then be used to determine inter-patient similarity. Patient attributes can first be aligned along a temporal dimension using significant temporal events such as admission to hospital, or the date of a first diagnosis or first treatment.

Once such sequences of temporal events have been arranged, the task is to identify patients who have similar attributes, that would allow a new patient’s likely trajectory in time to be inferred from the trajectories of past similar patients. Making such measures of similarity between two longitudinal sets of patient attributes is not a unique problem, and methods can borrow from techniques in information theory [16], time-series comparison [17], association indices in ecology and biological networks [18,19] and metric-learning [20].

Given a set of N patient attributes, a simple way to proceed is to create an N-dimensional vector assigning a value of 0 (not present) or 1 (present) for each attribute. The distance, or alternatively the similarity, between two sets of binary values can then be computed using a variety of measures, such as the Jaccard distance (as demonstrated in [19]). A more sophisticated approach might involve for example computing a weighted Euclidean distance in a transformed feature space [20], or weighting attributes based on their relative effect on patient outcomes.

**Optimal cohort selection**
For each index patient that we wish to make a treatment decision for, records of past patients can be queried to form an ordered set of virtual cohorts of similar patients. At the most stringent similarity level, an “exact match cohort” would include those patients who most closely match the index patient, while at the weakest level, a ‘full cohort’ includes all patients who simply satisfy an inclusion criteria.

Once inter-patient distance is computed, systematically increasing the cut-off distance from the index patient generates an ordered set of cohorts. Different distance measures can be compared, by looking at how cohort size and composition vary with cut-off distance. Methods such as agglomerative clustering [21] can also be used to grow larger cohorts. In this method, a distance between two cohorts is defined, for example, as the average of the distances between all patients in the cohorts and, in each successive iteration, the closest pair of cohorts is merged.

An alternative approach for growing larger cohorts from the index case, which does not require the use of an inter-patient distance measure, is to systematically relax attribute values. An attribute value is relaxed by extending the range of acceptably similar matches to adjacent values. For example, sex = male can be relaxed once to include females, while hypertension = moderate could be relaxed twice: a first time to include all degrees of hypertension, and a second time to also include normal blood pressure. In this approach, relaxation of one attribute value in the match cohort (level 0) will result in a new cohort. All such cohorts, each corresponding to a different attribute value being relaxed,
Commentary

Gallego, Walter, Day et al.

Commentary

match similarity measure that compares differences on criteria of Lifelines2 by introducing a Match & Mis similarity measures. Similan Cruiser as a normalized measure of R the cohort for which a selected measure of model fit such approach to selecting the most robust cohort is to pick the cohort for which a selected measure of model fit such as a normalized measure of R², reaches a maximum, and which then decreases as cohorts become larger.

Cohort visualization

Many EHR systems today provide capabilities to explore and visualize stored information [22]. Lifelines2 [23,24] aligns patient records by chosen events and can group patients by matching exact sequences of events. More sophisticated temporal queries (including temporal constraints and time spans between query elements) have been implemented in PatternFinder [25]. Visitors [26] support queries using higher abstractions based on clinical knowledge and nontemporal patient variables. Other software tools such as Outflow [27] visualize temporal trajectories of patients in preselected virtual cohorts. In particular, Caregiver [28] and CareCruiser [29] focus on visualizing clinical treatments and their effect on patients over time.

More complex algorithms cluster patients using similarity measures. Similan [30] relaxes the matching criteria of Lifelines2 by introducing a Match & Mismatch similarity measure that compares differences on temporal categorical sequences. MITHRA [31] measures similarity via a locally supervised metric learning, which accounts for the relative clinical relevance of patient measurements. DICON [32] conducts a cluster analysis based on key patient attributes and displays statistics for each cohort of similar patients.

Some systems (e.g., Gravi ++ [33] and DICON [32]) support visualization of patient cohorts as snapshots in a multidimensional space, helping clinicians understand a cohort’s composition. Initially, cohort visualization was used informally to assist decision makers. More recently, however, systems are being developed that combine more sophisticated analytics with cohort visualization and refinement (e.g., MITHRA, [31,34]).

Confounder control

The main challenge faced by cohort studies is treatment selection bias, where factors influencing treatment choice also independently influence patient outcome. In clinical trials, balance among treatment groups is achieved by randomization, and confounders are also minimized by stringent enrolment criteria, often resulting in patients with comorbidities being excluded. If observational studies fail to account for such confounders they may lead to biased estimates.

Strategies to deal with measured confounders include restriction, stratification, matching, inverse probability weighting and covariate adjustment. The study population should exclude patients for whom there is no treatment choice (e.g., due to contra-indications or on-going treatment [35]). Subgroup analysis via stratification is useful when there is suspicion that specific patient groups will react differently to the treatments under study, as well as to investigate the consistency and sensitivity of treatment effectiveness. Other preanalysis approaches include matching, often in fixed ratios, and inverse probability weighting, which creates a ‘standardized population’ by weighting individuals relative to their inverse probability of being included. Alternatively, parametric approaches like generalized linear models allow the incorporation of covariates without categorization and of specific temporal or nonlinear effects.

Lastly, it has become possible to estimate the probability of receiving treatment conditional on measured covariates (know as a propensity score). By stratifying, matching, weighting or adjusting using covariates or propensity score, treatment bias related to measured confounders can be removed [36]. Propensity scores can also guide exclusion criteria by removing low-score patients [37]. Analyses on each cohort can be performed adjusting for the potential confounding introduced by the unmatched attributes using propensity scores. These propensity scores can be estimated using automated methods like high dimensional propensity score [38].

Instrumental variables have been proposed as a means to emulate random treatment allocation so as to provide unbiased treatment effects, even in the presence of unmeasured confounders [39]. An instrumental variable is a factor that has a causal effect on treatment selection but that, conditional on treatment, neither has an effect on outcome, nor does it share any causes with it. Practical problems with this approach include the difficulty of finding suitable instruments (particularly in an automated fashion) and the inability to empirically verify whether such variables are unrelated to outcome other than through treatment.
Internal validation can be assessed (in real time) by measuring the robustness of the findings in relation to changes in adjustment methodologies, or by looking at variations in outcomes that are known to be unrelated to treatment.

Integrating results with existing evidence
It is important that results from point-of-care EHR-based cohort studies are not looked at in isolation but that they are presented to the clinician in the context of existing clinical guidelines and, if available, evidence from clinical trials. This integration of information serves two purposes: To validate the observational results obtained from clinical practice data and to identify and fill the inferential gaps that would otherwise have to be bridged unaided by data [40].

Advances in software technology and natural language processing are increasingly allowing for automated querying and meaningful visualization of evidence from published clinical trials [41]. Such a querying capability could be added to EHR systems to allow direct comparison of practice-based evidence and evidence-based practice by extracting results from similar inclusion criteria. Whenever an index patient does not fit the inclusion criteria of existing evidence, the gap can be filled using information from the EHR.

Discussion, conclusion & recommendations
It is important for the development of a learning health system to evolve new methods to support real-time decision-making at the point-of-care, to understand when these methods may provide valid results and to validate and integrate these findings with those from RCTs. Recently, EHR systems have been used to support RCTs by facilitating recruitment and follow-up [42]. If successful, evidence from these new types of RCTs could directly be integrated in EHR systems. Although they represent a great improvement from the current way clinical trials are performed, these point-of-care trials will still be insufficient to address the evidence gaps created by ethical or other limitations, when randomization is not possible [43].

Understanding the statistical and clinical significance of findings from studies using EHRs is a central problem. Significance levels may be adjusted depending on the treatments under consideration. For example, a more stringent significance level may be chosen before suggesting an invasive treatment with known risks. By doing so, a larger sample size is required to guarantee acceptable power. In some situations it may be necessary to filter the search by date (e.g., recent interventions); or by location or clinical setting (e.g., rural hospitals or specific institutions). On the other hand, combining data from various clinical settings increases the generalizability of findings and provide larger datasets, needed to look at rare conditions or events [44]. Distributed data networks (see list at [45]) with appropriate privacy algorithms, standards for defining data structures and software tools that can read across formats, are a step in this direction [46,47]. Clustering by provider or clinical setting can be integrated in the adjustment methodology.

An appropriate clinical decision would be supported by a statistically significant positive result that makes clinical sense, does not conflict with clinical guidelines and agrees with a patient’s values and expectations. Statistically significant results against intervention or null results can help reduce the tendency to act in order to avoid the regret of omission rather than commission. Results that are inconclusive, whether due to large variance or weak effect size are still valuable as additional evidence. Real-time interrogation of EHRs via virtual cohorts will offer more personalized care by complementing existing evidence, clinical guidelines and clinician experience.

Future perspective
Today, it is already possible to automate observational studies using information contained in the EHR. As the accuracy, consistency and completeness of EHRs improve so will the validity of these analyses. Advances in text processing will eventually allow for automated querying and meaningful visualization of related evidence from published clinical trials. If these capabilities are combined, both clinicians and patients will be able to visualize the effects of selected treatments on similar patients at the press of a button.

Acknowledgements
The authors would like to thank YS Low from the Stanford Center for Biomedical Informatics Research for expert feedback.

Financial & competing interests disclosure
This work was funded by National Health and Medical Research Council (NHMRC) Program Grant 568612, and Project Grants 1045548 and 1045065. Its contents are the responsibility of the authors and their institutions and do not reflect the views of the NHMRC. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

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

Open access
This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/
Executive summary

- Electronic health records (EHRs) from past patients are a source of information, which reflects patient’s treatment choices and their effects as they happen in actual clinical practice.
- This source of information is readily available and can be queried at the point-of-care to aid in decision-making for individual patients.
- Real-time querying tailored to individual patients requires: EHR-based phenotyping; quantifying inter-patient similarity; optimal cohort selection; cohort visualization; automated confounder control and integrating results with clinical guidelines and existing evidence from clinical trials.
- Real-time querying also requires real-time validation, an important open area of research. While bias from measured confounders can be minimized using automated propensity scores techniques, bias from unknown or unmeasured confounders can still threaten the validity of results. Evaluating results from cohorts with known estimates can increase confidence in these methods.
- Results from point-of-care EHR-based cohort studies should not be looked at in isolation but be presented in the context of existing clinical guidelines and any available evidence from clinical trials.

References

Bringing cohort studies to the bedside: framework for a ‘green button’ to support clinical decision-making

**Commentary**


44. Observational Health Data Sciences and Informatics (2014). www.ohdsi.org


### Basic Description

<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Journal of Comparative Effectiveness Research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ISSN</strong></td>
<td>2042-6305</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>Future Medicine Ltd.</td>
</tr>
<tr>
<td><strong>Country</strong></td>
<td>United Kingdom</td>
</tr>
<tr>
<td><strong>Status</strong></td>
<td>Active</td>
</tr>
<tr>
<td><strong>Start Year</strong></td>
<td>2012</td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td>Bi-monthly</td>
</tr>
<tr>
<td><strong>Language of Text</strong></td>
<td>Text in: English</td>
</tr>
<tr>
<td><strong>Refereed</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Abstracted / Indexed</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Serial Type</strong></td>
<td>Journal</td>
</tr>
<tr>
<td><strong>Content Type</strong></td>
<td>Academic / Scholarly</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>Print</td>
</tr>
<tr>
<td><strong>Website</strong></td>
<td><a href="http://www.futuremedicine.com/loi/cer">http://www.futuremedicine.com/loi/cer</a></td>
</tr>
</tbody>
</table>

**Description:** Provides a forum for all stakeholders, featuring comparative effectiveness studies, methodology papers, review articles, policy papers, perspective pieces.

### Subject Classifications

### Additional Title Details

### Publisher & Ordering Details

### Online Availability

### Abstracting & Indexing

### Demographics