Are electronic health records ready for clinical trial use?

Lucy Plumb and Dorothea Nitsch

Standfirst

Routinely collected data from electronic health records (EHRs) could potentially be used to facilitate clinical trials. Use of computational phenotypes shows promise for detecting trial-eligible paediatric patients; however, moving from EHR identification to recruitment requires consideration of legal, ethical and logistical challenges.

Refers to Denburg, M.R. et al. Using electronic health record data to rapidly identify children with glomerular disease for clinical research. *J. Am. Soc. Nephrol.* **30**, 2427–2435 (2019).

Commentary

For paediatric nephrologists, who are used to dealing with rarities, glomerular disorders can be considered one of the more 'common' manifestations of kidney disease in children and young people. Nephrotic syndrome, which affects 2–7 per 100,000 children[1], might be one of the first nephrological encounters for many paediatricians in training. Glomerular disease also accounts for a considerable proportion of long-term kidney injury: 5–14% of all cases of chronic kidney disease and 15–29% of cases of end-stage renal disease worldwide[2]. Despite its relative frequency, therapeutic advances in paediatric glomerular disease have not been forthcoming, in part because small patient numbers limit the achievable statistical power of clinical trials. For example, a sample size of 348 patients would be needed to detect a minimum 20% difference in the protein:creatinine ratio from a mean baseline value of 150mg/mmol in two compared treatment groups. As such, it is of no surprise that children are often recruited alongside adults in therapeutic trials, although it is unclear how generalizable results are to the paediatric population. One systematic review identified only 27 paediatric nephrology clinical trials, which accounted for 2.6% of all nephrology studies and 0.9% of all studies specifically in paediatric populations[3].Clearly, adequately powered clinical trials are required to determine the most effective strategies to prevent disease progression in children, not least for those with glomerular disease. But how can we identify sufficient numbers of eligible participants in an efficient and cost-effective manner?

In the *Journal of the American Society of Nephrology*, Denburg and colleagues offer a solution to this question that has been under our noses for some time: the (almost) ubiquitous electronic health record (EHR)[4]. In their article, Denburg *et al*. described the derivation of a computational phenotype

from data elements identified in the medical records and clinical encounters of 231 paediatric patients with confirmed glomerular disease. An iterative approach was undertaken to test the rule-based algorithm using blinded, structured chart reviews in a single tertiary hospital. These findings were subsequently validated in another eight specialist. During this period of refinement, the team modified the algorithm requirements to account for codes linked to falsely positive cases, such as 'acute glomerulonephritis' and 'glomerulosclerosis'. The resulting computational phenotype, which comprises clinical encounters (that is, three or more nephrology reviews), diagnostic and kidney biopsy procedure codes, has a high overall sensitivity (96%), specificity (93%) and negative predictive value (97%) for detecting children with glomerular disease. Unfortunately, the data published in the paper are not entirely transparent; the total patient numbers used for validation quoted in the abstract do not match those given in the tables. Nevertheless, in theory, the use of robust computational phenotypes to identify disease-specific paediatric cohorts could accelerate the identification of potential participants for clinical trials.

However, EHRs are designed primarily to record clinical interactions and to audit and improve health-care provision, not to recruit clinical trial participants, which has a number of implications. First, there is a legal dimension. For example, under the terms of the European General Data Protection Regulation, patients identified by the algorithm in this research cannot be directly approached by trial investigators without having previously given consent to being contacted regarding trial recruitment at some time in the future. Patients must also be offered the opportunity to opt out of such data use[5]. These issues can be circumvented by, for example, implementing alerts in EHR systems that first prompt treating clinicians to obtain such consent and subsequently to recruit eligible patients when they are next seen.

Second, we must be mindful that for any diagnostic algorithm, the trade-off between sensitivity and specificity necessarily results in both false positive and false negative findings. Approaching a patient who has been falsely identified as eligible for recruitment could result in increased anxiety for the child and their family, additional tests and, in the worst-case scenario, exposure to an unnecessary treatment. Moreover, detection of a potentially trial-eligible patient may not automatically equate to the patient being fully informed about their condition. Clearly, rigorous checks in close collaboration with treating physicians will be required for confirmation of a patient's eligibility for recruitment before patients are approached.

Third, any algorithm can only be used in care settings similar to those in which the algorithm was originally developed. For centres using substantially different care pathways and protocols, the restriction to three or more clinical encounters might lead to false-negative findings (for example, if patients are hospitalized for prolonged periods). Issues around billing and/or funding incentives may also lead to particular codes being used more often in some settings than in others. A requirement for attending a nephrologist on three or more occasions, as included in this study's algorithm, may mean that populations that do not seek (or do not attend) health-care appointments, which may include disadvantaged children, are under-represented in clinical trials. Although children from low-income families benefit from government-funded health-care initiatives such as Medicaid or the Children's Health Initiative Programme (CHIP), their eligibility criteria and the benefits offered vary by jurisdiction, which may also affect the phenotype's generalizability across regions.

Fourth, even a well-designed and validated digital phenotype is only as good as the data captured within the EHR system. For example, if a particular disease type is associated with having a particular ethnicity, the false-positive rate might also vary by ethnicity[6]. Data capture of important confounding variables (such as ethnicity or measures of socioeconomic deprivation) and high rates of missing data and/or misclassification are ongoing challenges for EHR-based research[7].

In summary, Denburg and colleagues have demonstrated that it is possible to use EHRs to identify, with a high degree of accuracy, patient populations for whom pragmatic clinical trials are needed. Moving on from identification to recruitment requires consideration of several legal, ethical and logistical issues; working with patients and their families, health-care professionals, clinical trial investigators and data providers is necessary to ensure that the solutions to these issues are acceptable to patients, their families and the wider public.

Lucy Plumb[1,2]* and Dorothea Nitsch[1,3]

[1]UK Renal Registry, Bristol, UK.
[2]Population Health Sciences Institute, University of Bristol Medical School, Bristol, UK.
[3]Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.

*e-mail: dorothea.nitsch@lsthtm.ac.uk

References

1. Rheault MN & Wenderfer SE. Evolving epidemiology of pediatric glomerular disease. *Clin. J. Am. Soc. Nephrol*. 13, 977-978 (2018).
2. Harambat J, Van Stralen KJ, Kim JJ, Tizard EJ. Epidemiology of chronic kidney disease in children. Pediatric nephrology. 2012 Mar 1;27(3):363-73.
3. Inrig JK *et al*. The landscape of clinical trials in nephrology: a systematic review of Clinicaltrials. gov. *Am. J. Kidney Dis.* **63**, 771-780 (2014).
4. Denburg MR *et al*. Using electronic health record data to rapidly identify children with glomerular disease for clinical research. *J. Am. Soc. Nephrol*. **30**, 2427-2435 (2019).
5. Raman SR *et al*. Leveraging electronic health records for clinical research. *Am. Heart J.* **202**, 13-19 (2018).
6. Perneger TV, Whelton PK, Klag MJ, Rossiter KA. Diagnosis of hypertensive end-stage renal disease: effect of patient's race. *Am. J. Epidemiol.* **141**, 10-15 (1995).
7. Horth RZ *et al*. Use of electronic health records from a statewide health information exchange to support public health surveillance of diabetes and hypertension. *BMC Public Health* **19**, 1106 (2019).

Acknowledgements

L.P. is supported by the National Institute for Health Research (NIHR Doctoral Research Fellowship, DRF-2016-09-055). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The funding body had no role in the design, collection, analysis, or interpretation of data in the writing of the manuscript, or in the decision to submit the manuscript for publication. D.N. is co-applicant on two GSK-funded, open-label studies in sub-Saharan Africa that are unrelated to this work.

Competing interests

The authors declare no competing interests.