Published in final edited form as:

J Pediatr Urol. 2018 October ; 14(5): 374–381. doi:10.1016/j.jpuro.2018.04.033.

“Minimally Invasive Research?” Use of the Electronic Health Record to Facilitate Research in Pediatric Urology

VM Vemulakonda^a, RA Bush^{b,c}, and MG Kahn^{d,e}

^aDepartment of Pediatric Urology, Children’s Hospital Colorado; Division of Urology, Department of Surgery, University of Colorado Denver Anschutz Medical Campus, Aurora, Colorado, USA

^bClinical Informatics, Rady Children’s Hospital San Diego

^cUniversity of San Diego Beyster Institute for Nursing Research, San Diego, California, USA

^dDepartment of Pediatrics, Colorado Clinical and Translational Sciences Institute and Colorado Center for Personalized Medicine, University of Colorado Denver Anschutz Medical Campus

^eResearch Informatics, Children’s Hospital Colorado, Aurora, Colorado, USA

Abstract

Background—The electronic health record (EHR) was designed as a clinical and administrative tool to improve clinical patient care. Electronic health care systems have been successfully adopted across the world through use of government mandates and incentives.

Methods—Using electronic health record, health information system, electronic medical record, health information systems, research, outcomes, pediatric, surgery, and urology as initial search terms, the literature focusing on clinical documentation data capture and the EHR as a potential resource for research related to clinical outcomes, quality improvement, and comparative effectiveness was reviewed. Relevant articles were supplemented by secondary review of article references as well as seminal articles in the field as identified by the senior author. Findings: U.S. federal funding agencies, including the Agency for Healthcare Research and Quality, the Patient-Centered Outcomes Research Institute, the National Institutes of Health and the Food and Drug Administration have recognized the EHR’s role to support research. The main approaches to using EHR data include enhanced lists, direct data extraction, structured data entry, and unstructured data entry. The EHR’s potential to facilitate research, overcoming cost and time burdens associated with traditional data collection, has not resulted in widespread use of EHR-based research tools.

Conclusion—There are strengths and weaknesses for all existing methodologies of using EHR data to support research. Collaboration is needed to identify the method that best suits the institution for incorporation of research-oriented data collection into routine pediatric urologic clinical practice.

CONFLICT OF INTEREST

The authors have no financial relationships relevant to this work.

ETHICAL APPROVAL

This work reflects a review of the current literature and does not fall under the purview of the Institutional Review Board.

Keywords

electronic health record; comparative effectiveness; surgical outcomes; health information technology

INTRODUCTION

Investment in Health Information Technology (HIT) was designed to improve health care quality, decrease costs, and enhance research to improve patient well-being. While electronic health record (EHR) adoption and use continues to grow across the world [1,2], its use in the U.S. has expanded clinical data capture, creating an enormous volume of electronically accessible patient data available for research in clinical outcomes, quality initiatives, and comparative effectiveness [3–5].

Research using EHR-based networks to impact outcomes is a funding priority for several U.S. funding agencies [6]. Methodologies such as comparative effectiveness research (CER) [7] and pragmatic trials [8] benefit from the volume of data and the ability to measure outcomes within a natural practice setting, examine health care delivery system effects, and involve heterogeneous populations [9]. Combining data from diverse real-world clinical settings rather than tightly constrained clinical trials also strengthens the generalizability of findings [7,8,10]. Compared to traditional data collection, use of EHR-based documentation tools may allow for more comprehensive data capture than administrative databases and more uniform data collection than manual data extraction. Additionally, EHR-based data collection may help to overcome the financial and time burdens associated with manual data collection [9,11].

Efficient clinical documentation maximizes patient throughput and billing while research documentation maximizes accuracy and completeness; these competing objectives lead to conflicting optimization strategies. Ideally, EHR documentation methods are flexible enough to adapt to different provider styles, thereby improving clinical efficiency; however, these variations may limit the ability to consistently collect data for research purposes. As health care providers are not trained to gather data for research, [12] clinical data captured through the EHR may be incomplete, inconsistent, and inaccurate for research analysis [3,4,12–14]. The purpose of this article is to compare existing EHR-based tools for research and to share lessons learned from incorporating research-oriented data collection tools into routine pediatric urologic practice (Table 1).

METHODS

The National Library of Medicine's PubMed was used to search for articles including the terms: pediatric, surgery, urology, electronic health record, electronic medical record, EHR, EMR, outcomes, health information system, and research from any date until November, 2017. Given the paucity of pediatric citations, the search was expanded to adults and the phrase comparative effectiveness research was added. The abstracts of all identified articles were reviewed by two authors (VMV and RAB). Full text of articles relevant to the use of the EHR in pediatric outcomes research were reviewed. Additional articles were identified

via review of the references of articles deemed relevant on primary search. Data regarding the application of the EHR for research as well as associated strengths and weaknesses were extracted for incorporation into the review article. Areas where extracted data were limited were augmented by additional landmark articles in the field of health informatics as identified by the senior author (MGK).

USE OF EHR-GENERATED ENHANCED LISTS TO IDENTIFY PATIENTS

Approach

Using the EHR to create an enriched list of patients who meet criteria such as age, ICD9/ICD10 codes, and encounter dates is one approach to identify eligible patients for subsequent traditional data abstraction [15]. If custom templates or checklists are available in the EHR, these data elements may be used to improve identification of patients and augment data abstracted through manual review [14]. This method is scalable across multiple sites.

Example

Pediatric urologists at three practices were able to evaluate imaging practice patterns of infants with hydronephrosis identified by ICD-9 code, with data extraction performed using EHR reporting tools for patient demographics and orders and manual chart review for imaging findings [16]. A similar approach has been used by the Pediatric Emergency Care Applied Research Network, which has used chief complaint groupings to trigger use of a standardized trauma data collection tool [17].

Considerations

Advantages of this hybrid method include: ease in accessibility as patient lists do not require sophisticated informatics personnel and increased efficiency of patient identification via standardized inclusion criteria while maintaining detailed information capture via traditional chart review. Augmenting EHR-based data capture with chart review also permits capture of data stored as narrative text in clinical notes, discharge summaries, and operative reports, that may not be accessible using electronic reporting tools [14].

The biggest limitation to this method is that it is only the first step in data extraction. It must often be supplemented by other data collection methods, such as structured data elements in the EHR or traditional chart review. Chart review is frequently conducted by individuals with extensive training and familiarity with local medical documentation practices to increase reliability and minimize abstractor variation during the multiple steps of data extraction. After extraction, data are reviewed to ensure quality, adding to the time and cost of this approach [11,18]. These potential costs may be even more significant when abstracting surgical data. In the absence of standardized charting, operative details may be difficult for research personnel to identify, requiring surgeons to collect and review data for accuracy. Thus, enhanced patient lists alone may not be sufficient to address barriers associated with obtaining data from EHR systems.

ELECTRONIC DATA EXTRACTION USING PRE-EXISTING DATA FIELDS

Approach

This method extracts data from pre-existing EHR data fields include standardized data elements, such as demographics, orders, machine-generated lab results, flowsheets, vital signs, appointment schedules, and billing data [19].

Examples

In order to evaluate medical utilization patterns for children with autism spectrum disorder (ASD), four different queries extracting data from administrative and billing modules as well as patient problem lists and chief complaints, elucidated the need to use different databases to reflect different workflows used to capture patient diagnoses [20]. A multi-center study of pediatric urology clinic non-attendance, used existing data to examine return to clinic patterns associated with provider, time of year, and length to next appointment [21].

Considerations

The most easily accessible EHR data is extracted from variables collected in pre-existing data fields. Because these data fields are pre-built and embedded within EHR clinical functionality, individual researchers do not face the burden of developing and implementing data capture tools to collect patient information. Due to pre-defined and agreed upon terms, data are standardized a priori to facilitate research data collection across multiple specialties within an institution as well as among institutions [21]. Finally, data are fairly diverse, encompassing demographics, diagnoses, encounter visit information, procedures, medications, labs and radiographic studies ordered, and, in select cases, study results [12,22].

This approach, however, has several limitations. Differences in how data are identified may lead to differences in data extraction. The four different data queries extracting data for individuals with ASD from the same EHR system yielded substantially different results because of different workflows used to capture patient diagnoses and relied on inclusion of emergency department records to capture individuals without insurance coverage [20]. Issues of data variability can be exacerbated in multi-center research when participating institutions have different storage structures, which require system-specific extractions and data management. Moreover, regulatory and institutional limitations to data sharing may prohibit the extraction of certain data to ensure privacy, thereby further limiting data available for the pooled analysis [21]. Adoption of specialty-specific checklists may vary across sites. In a study assessing use of a standardized checklist for pathology results, up to 25% of reports did not include use of the checklist [23]. Data capture may also be incomplete due to pre-determined definitions. For example, in a study of vesicoureteral reflux practice patterns, available order data did not adequately capture outside imaging and lab studies, leading to under-estimation of physician use of these studies [24]. Use of data primarily intended for clinical and billing uses may also not adequately capture data needed for research. In the vesicoureteral reflux study, race was not documented in 7% of patients and encounter diagnosis did not accurately reflect laterality or associated reflux nephropathy in almost 20% of patients. These results are similar to prior studies in patients with heart

failure [25]. Similarly, in the clinic non-attendance study, identification of diagnosis in patients who did not attend clinic was limited since diagnoses are generally assigned during a patient encounter [21]. Finally, working with data collected for clinical uses rather than research may lack the temporal relationship needed to determine disease cause [14]. If diagnosis codes indicate a patient has urinary retention and a UTI, is the patient in retention because of the UTI or did the UTI result from the retention? These limitations highlight the need for well-tailored research questions, quality review of data captured to ensure accuracy, and augmentation of data captured from pre-existing fields with other methods to ensure adequate completeness of data to answer the research question.

STRUCTURED (DISCRETE) DATA ENTRY

Approach

All EHR systems permit creation of structured data entry systems, which are user-designed and developer-created electronic forms or templates that optimize data completeness and standardization [26]. Rather than documenting patient information using unstructured text, discrete data elements using check boxes or drop-down menus may be used to capture a wide variety of data, including patient complaints, imaging findings, or treatment plans. These templates can be modified to match encounter specific variables and existing structured reports so that anyone reviewing the notes sees information in a structured narrative format.

Methods used to capture structured data elements can vary (Figure 1). Data may be collected via structured data forms that are integrated into an unstructured note. Alternatively, individual data elements can be embedded within the exist chart, allowing for more customized documentation based on provider preference.

Examples

In a study of infants with hydronephrosis, individual data elements were developed to capture data on patient presentation, imaging findings, diagnosis, and treatment plan. These elements were then incorporated into pre-existing note templates or could be added to unstructured notes based on physician preference. The use of discrete data elements was found to be technically feasible with good physician adoption. Following the initial implementation of discrete data elements, there was good surgeon utilization (>80%) with accuracy comparable to manual chart review. Several factors supported the strong adoption. Providers were involved in the development phase to ensure “buy in” prior to implementation. Providers familiar with using a semi-structured note had higher rates of adoption than those using an unstructured note, leading to universal adoption of the semi-structured note in the practice. Adoption was improved by tracking adherence, with adherence improving over time [11]. Structured data forms have been successfully used for research data collection in the ImproveCareNow network, a multicenter registry studying pediatric inflammatory bowel disease [27].

Considerations

The integration of unstructured text with discrete data fields facilitates physician data capture, allows freedom of expression, supports sharing of standardized data among EHR systems, and provides structured data for meaningful use reporting, quality assurance, and clinical research [28]. Changes in clinical charting practices and workflows requires commitment from invested stakeholders on whom data collection relies. Processes such as visual cues to aid in identifying incomplete data collection may remind providers to complete structured data elements. Additionally, ongoing education of providers and assessment of adherence may improve completion. Unlike other forms of structured data, where individual elements may be deleted without impeding the ability to complete charting, structured forms often require completion of all elements within the form prior to integration into the note as a quality check to ensure completeness of documentation. While mandating data collection may help to improve data capture, this approach may not always be feasible, as patients may not provide all necessary data or may deviate from the underlying assumptions of the form, with unanswered fields impeding clinical charting and encounter completion [27].

This data collection and extraction approach allows for more customized data collection than other EHR-based approaches, but it may be less accessible due to the significant up-front investment in changing clinical documentation patterns and the potential impact of multiple additional data elements on clinician workflow [29]. There is an inherent tension between the limited expressiveness of structured data capture and the ability to extract unambiguous concepts; as expressiveness and flexibility increases, analytical usability decreases. Clinicians are reluctant to switch from natural prose to clinic documentation templates (“checklist medicine”) despite increased accuracy, reliability, and understandability associated with structured documentation [30].

Provider “buy in” and pre-implementation planning is even more critical for multi-center data collection. Providers and investigators may be unfamiliar with data extraction tools and may not utilize structured data elements even when available due to difficulty in changing existing clinic workflows [31]. Implementation of discrete data elements in the operative setting may also lead to delays in operative report generation, despite potential long-term improvements in data quality and accessibility [32].

These issues are exacerbated as data is collected in diverse environments and by different people. If the research is longitudinal, it is likely to incorporate several practice settings. The variety of input needed may be further increased if the population of interest is treated at both academic centers and community locations or across both urban and rural settings. As a result, the sheer volume and diversity of discrete data elements needed to adequately capture data may create an insurmountable obstacle to effective research.

The development and implementation of discrete data elements requires a dedicated and experienced informatics team to construct the data collection/clinical documentation forms as well as a sophisticated medical informatics team to aid in data extraction. As a result, the technical and organizational infrastructure needed to utilize this approach may limit its use

to programs with well-established informatics departments, thereby limiting the ability to foster large- scale, representative, collaborative research networks.

NATURAL LANGUAGE PROCESSING

Approach

Existing clinical documentation workflows often rely on unstructured text [31]. In response, informaticists have developed natural language processing (NLP), which uses computer algorithms to identify and extract structured data elements from unstructured text (often called “free text”). Ideally, NLP can convert providers’ narratives into discrete variables, which can then be stored in structured databases [33]. For example, information extraction, a common NLP method, has been used in EHR-based studies to enable data extraction, allowing continue unlimited expressiveness among clinicians without impacting existing clinical work flows [14]. To date, no examples of this approach were found in the pediatric urology literature.

Considerations

While the upfront costs of NLP are minimal, the resources needed to produce an analysis-ready database can be extensive. The quality and availability of text is often widely variable, with non-standard abbreviations and variant spellings of similar words. This text variability challenges the creation of an effective NLP program [10]. Very small changes in documentation or dictation styles can dramatically alter the ability to identify relevant text. Experts with specific expertise in NLP and programmers with extensive training are needed to construct the computer algorithm, complete extensive validation, and ensure stable performance of the extraction program. To overcome some of these limitations, data may be “pre-processed” by using spell-checking, word sense disambiguation to identify the correct meaning of words with multiple meanings, and marking words as a particular part of speech (nouns versus verbs versus adjectives) [34].

Even with adequate support, the ability to extract data across multiple configurations of text (unstructured text, dictated text, and structured templates) may limited [35]. Furthermore, clinical text may have cut and pasted characters that are not classified as text and may therefore not be captured. NLP is necessarily retrospective and may be limited in capturing temporal relationships. It has been most sensitive when records are screened by enhanced patient lists or other techniques [34,36]. Because NLP programs are tuned to detect specific provider or departments’ idiosyncrasies, they are not generalizable to other research applications or scalable across practice settings.

DISCUSSION

While the EHR has the potential to serve as a valuable data resource for clinical data, there are always costs and tradeoffs to be made to data gathering approaches because the EHR does not have pre-built data capture tools built into workflows needed for comprehensive clinical research. Planning is critical and needs to include an assessment of research goals, resources, available examples, existing templates for use or reconfiguration, and potential approaches to supplement data gathering. Working with colleagues and other sites to

establish a common understanding of research concepts, establish surgeon “buy in” to alterations to workflow, and ensure compatible data are captured prior to wide-scale implementation facilitate effective collaboration and may help to minimize the risk of unanticipated issues once data collection has begun.

Collecting research data without compromising clinician’s commitment to patient care is a promising step toward decreasing research costs, increasing patient-centered research, and speeding the rate of new medical discoveries. The potential of EHR-based research is best realized through the development and optimization of discrete data elements to support clinical workflow and research data capture. To facilitate the utility of these discrete data elements, it is essential they be incorporated into existing clinical workflows with minimal disruption of provider documentation practices. With this goal in mind, four general steps have been proposed to deliver a complete, accurate, and usable discrete data element approach: 1) Establish a clinical advisory committee, including champions of EHR-supported research, to identify questions of high clinical impact and patient importance; 2) Identify “deal breakers” for structured data entry with specific attention to physician resistance; 3) Identify workflows to facilitate high-quality, complete data entry; and 4) Identify technology platforms needed for seamless integration.[37]

Engaging specialty societies to identify meaningful and impactful concepts that would be supported by multiple practitioners willing to cede some individuality for access to a comprehensive, multi- center data warehouse. More specifically, specialty societies could propose and endorse standard datasets to encourage adoption by their members [38] to be incorporated into existing repositories of structured data elements, such as the NIH common data element (CDE) repository and to standardize structured data elements that may be shared across multiple practices and multiple EHR systems.

CONCLUSIONS

EHR adoption has led to improvements in clinical documentation and patient outcomes [39,40]. The EHR can provide the needed existing data for successful outcomes and clinical effectiveness research, depending on the structure of the research question. However, much like other large-scale data sources, the quality of the data obtained from the EHR is reliant on upfront buy-in from physicians and institutional stakeholders as well as the availability of resources to support nuanced data collection and extraction. As a result, development of research questions that rely upon EHR-based data must be tailored to fit the clinical data available. Despite the steep upfront costs of developing and implementing comprehensive research collection tools that do not impede clinical workflow, the EHR offers a promising alternative to traditional data resources in developing targeted collaborative data networks that allow for more definitive understanding of practice patterns and patient outcomes in real-world pediatric urologic practice.

Acknowledgments

FUNDING SOURCE

This project was supported in part by: the Agency for Healthcare Research and Quality (grant numbers R00 HS022404 (RAB) and K08 HS024597-01 (VMV) as well as by the American Urological Association Rising Stars in Urology Research Award Program and the Frank and Marion Hinman Urology Research Fund (VMV). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality, the American Urological Association, or the Frank and Marion Hinman Urology Research Fund.

References

1. Jha AK, Doolan D, Grandt D, Scott T, Bates DW. The use of health information technology in seven nations. *Int J Med Inform* [Internet]. 2008 Dec; 77(12):848–54. [cited 2018 Jan 23] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18657471>.
2. Anderson JG, Vagnoni E. Social, ethical and legal barriers to e- health. *Int J Med Inform* [Internet]. 2017 Mar; 76(5–6):480–3. [cited 2018 Jan 23] Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0959652617306297>.
3. Embi PJ, Hebert C, Gordillo G, Kelleher K, Payne PRO. Knowledge management and informatics considerations for comparative effectiveness research: a case-driven exploration. *Med Care* [Internet]. 2013 Aug; 51(8 Suppl 3):S38–44. [cited 2013 Nov 25] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23793050>.
4. Danciu I, Cowan JD, Basford M, Wang X, Saip A, Osgood S, Shirey- Rice J, Kirby J, Harris PA. Secondary use of clinical data: The Vanderbilt approach. *J Biomed Inform* [Internet]. 2014 Dec. 52:28–35. [cited 2017 Dec 20] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24534443>.
5. Brown JS, Kahn M, Toh S. Data quality assessment for comparative effectiveness research in distributed data networks. *Med Care* [Internet]. 2013 Aug; 51(8 Suppl 3):S22–9. [cited 2013 Nov 29] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23793049>.
6. Wilcox A, Randhawa G, Embi P, Cao H, Kuperman GJ. Sustainability considerations for health research and analytic data infrastructures. *EGEMS (Washington, DC)* [Internet]. 2014; 2(2):1113. [cited 2017 Dec 21] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25848610>.
7. Tunis SR, Benner J, McClellan M. Comparative effectiveness research: Policy context, methods development and research infrastructure. *Stat Med* [Internet]. 2010 Aug 30; 29(19):1963–76. [cited 2012 Nov 6] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20564311>.
8. Richesson RL, Green BB, Laws R, Puro J, Kahn MG, Bauck A, Smerek M, Van Eaton EG, Zozus M, Ed Hammond W, Stephens KA, Simon GE. Pragmatic (trial) informatics: a perspective from the NIH Health Care Systems Research Collaboratory. *J Am Med Informatics Assoc* [Internet]. 2017 Sep 1; 24(5):996–1001. [cited 2017 Dec 21] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28340241>.
9. Chang DC, Rhee DS, Papandria D, Aspelund G, Cowles RA, Huang EY, Chen C, Middlesworth W, Arca MJ, Abdullah F. Outcomes research in pediatric surgery. Part 2: how to structure a research question. *J Pediatr Surg* [Internet]. 2011 Jan; 46(1):226–31. [cited 2012 Nov 4] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21238673>.
10. Raebel MA, Haynes K, Woodworth TS, Saylor G, Cavagnaro E, Coughlin KO, Curtis LH, Weiner MG, Archdeacon P, Brown JS. Electronic clinical laboratory test results data tables: lessons from Mini-Sentinel. *Pharmacoepidemiol Drug Saf* [Internet]. 2014 Jun; 23(6):609–18. [cited 2018 Jan 11] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24677577>.
11. Chalmers DJ, Deakne SJ, Payan ML, Torok MR, Kahn MG, Vemulakonda VM. Feasibility of Integrating Research Data Collection into Routine Clinical Practice Using the Electronic Health Record. *J Urol* [Internet]. 2014 Oct; 192(4):1215–20. [cited 2017 Dec 21] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24793730>.
12. Bush RA, Kuelbs C, Ryu J, Jiang W, Chiang G. Structured Data Entry in the Electronic Medical Record: Perspectives of Pediatric Specialty Physicians and Surgeons. *J Med Syst* [Internet]. 2017 May 21; 41(5):75. [cited 2017 Apr 18] Available from: <http://link.springer.com/10.1007/s10916-017-0716-5>.
13. Curcin V, Soljak M, Majeed A. Managing and exploiting routinely collected NHS data for research. *Inform Prim Care* [Internet]. 2012 Jan; 20(4):225–31. [cited 2014 Jan 30] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23890333>.

14. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PRO, Bernstam EV, Lehmann HP, Hripcsak G, Hartzog TH, Cimino JJ, Saltz JH. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care* [Internet]. 2013 Aug; 51(8 Suppl 3):S30–7. [cited 2017 Dec 20] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23774517>.
15. Pakhomov SVS, Finley G, McEwan R, Wang Y, Melton GB. Corpus domain effects on distributional semantic modeling of medical terms. *Bioinformatics* [Internet]. 2016 Aug 16;32(23):btw529. [cited 2017 Dec 21] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27531100>.
16. Vemulakonda VM, Chiang G, Corbett ST. Variability in Use of Voiding Cystourethrogram During Initial Evaluation of Infants With Congenital Hydronephrosis. *Urology* [Internet]. 2014 May; 83(5):1135–8. [cited 2017 Dec 21] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24397941>.
17. Deakynne SJ, Bajaj L, Hoffman J, Alessandrini E, Ballard DW, Norris R, Tzimenatos L, Swietlik M, Tham E, Grundmeier RW, Kuppermann N, Dayan PS. Pediatric Emergency Care Applied Research Network (PECARN). Development, Evaluation and Implementation of Chief Complaint Groupings to Activate Data Collection. *Appl Clin Inform* [Internet]. 2015; 6(3):521–35. [cited 2018 Jan 3] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26448796>.
18. Walther B, Hossin S, Townend J, Abernethy N, Parker D, Jeffries D, Biondi-Zoccai G, editor [cited 2017 Aug 10] Comparison of Electronic Data Capture (EDC) with the Standard Data Capture Method for Clinical Trial Data; *PLoS One* [Internet]. 2011. Sep 23, e25348 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21966505>
19. West SL, Johnson W, Visscher W, Kluckman M, Qin Y, Larsen A. The challenges of linking health insurer claims with electronic medical records. *Health Informatics J* [Internet]. 2014 Mar; 20(1): 22–34. [cited 2016 Jan 14] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24550563>.
20. Bush RA, Connelly CD, Pérez A, Barlow H, Chiang GJ. Extracting autism spectrum disorder data from the electronic health record. *Appl Clin Inform* [Internet]. 2017; 8(3):731–41. [cited 2017 Jul 19] Available from: <http://www.schattauer.de/index.php?id=1214&doi=10.4338/ACI-2017-02-RA-0029>.
21. Bush RA, Vemulakonda VM, Corbett ST, Chiang GJ. Can we predict a national profile of non-attendance paediatric urology patients: a multi-institutional electronic health record study. *Inf Prim Care*. 2014; 21(3):132–8. [2014/09/11]
22. Hassell LA, Parwani AV, Weiss L, Jones MA, Ye J. Challenges and opportunities in the adoption of College of American Pathologists checklists in electronic format: perspectives and experience of Reporting Pathology Protocols Project (RPP2) participant laboratories. *Arch Pathol Lab Med* [Internet]. 2010 Aug; 134(8):1152–9. [cited 2017 Dec 21] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20670135>.
23. Hassell L, Aldinger W, Moody C, Winters S, Gerlach K, Schwenn M, Perriello D. Electronic capture and communication of synoptic cancer data elements from pathology reports: results of the Reporting Pathology Protocols 2 (RPP2) project. *J Registry Manag* [Internet]. 2009; 36(4) [cited 2017 Dec 21] 117–24–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20795553>.
24. Vemulakonda VM, Janzen N, Sevick C, Chiang G. Feasibility of establishing a multi-center research network. *South Central Section American Urologic Association*. 2017
25. Baker K, Dunwoodie E, Jones RG, Newsham A, Johnson O, Price CP, Wolstenholme J, Leal J, McGinley P, Twelves C, Hall G. Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy. *Int J Med Inform* [Internet]. 2017 Jul. 103:32–41. [cited 2017 Oct 16] Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1386505617300722>.
26. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *J Am Med Inform Assoc* [Internet]. Jan; 13(3):277–88. [cited 2014 May 27] Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1513664&tool=pmcentrez&rendertype=abstract>.
27. Marsolo K, Margolis PA, Forrest CB, Colletti RB, Hutton JJ. A Digital Architecture for a Network-Based Learning Health System – Integrating Chronic Care Management, Quality Improvement, and Research. *eGEMs (Generating Evid Methods to Improv patient outcomes)* [Internet]. 2015

- Aug 17.3(1):16. [cited 2018 Apr 9] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26357665>.
28. Johnson SB, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, Bright T, Van Vleck T, Wrenn J, Stetson P. An electronic health record based on structured narrative. *J Am Med Inform Assoc* [Internet]. Jan; 15(1):54–64. [cited 2014 May 27] Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2274868&tool=pmcentrez&rendertype=abstract>.
 29. Leu MG, O'Connor KG, Marshall R, Price DT, Klein JD. Pediatricians' use of health information technology: a national survey. *Pediatrics* [Internet]. 2012 Dec; 130(6):e1441–6. [cited 2014 Aug 8] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23166335>.
 30. Marill KA, Gauharou ES, Nelson BK, Peterson MA, Curtis RL, Gonzalez MR. Prospective, randomized trial of template-assisted versus undirected written recording of physician records in the emergency department. *Ann Emerg Med* [Internet]. 1999 May; 33(5):500–9. [cited 2014 Jun 2] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10216325>.
 31. Katzan I, Speck M, Dopler C, Urchek J, Bielawski K, Dunphy C, Jehi L, Bae C, Parchman A. The Knowledge Program: an innovative, comprehensive electronic data capture system and warehouse. AMIA. Annu Symp proceedings AMIA Symp [Internet]. 2011; 2011:683–92. [cited 2017 Aug 10] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22195124>.
 32. Robinson JR, Huth H, Jackson GP. Review of information technology for surgical patient care. *J Surg Res* [Internet]. 2016 Jun 1; 203(1):121–39. [cited 2017 Aug 16] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27338543>.
 33. Townsend Hilary. Natural Language Processing and Clinical Outcomes: The Promise and Progress of NLP for Improved Care. *J AHIMA* [Internet]. 2013; 84(2):44–5. [cited 2017 Dec 21] Available from: <http://bok.ahima.org/doc?oid=106198#.WjxoIIWnHZ4>.
 34. Grundmeier RW, Masino AJ, Casper TC, Dean JM, Bell J, Enriquez R, Deakyne S, Chamberlain JM, Alpern ER. Pediatric Emergency Care Applied Research Network. Identification of Long Bone Fractures in Radiology Reports Using Natural Language Processing to support Healthcare Quality Improvement. *Appl Clin Inform* [Internet]. 2016 Nov 9; 7(4):1051–68. [cited 2017 Aug 16] Available from: <http://www.schattauer.de/index.php?id=1214&doi=10.4338/ACI-2016-08-RA-0129>.
 35. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. [cited 2017 Aug 10] Extracting information from textual documents in the electronic health record: a review of recent research; *Yearb Med Inform* [Internet]. 2008. 128–44. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18660887>.
 36. Doan S, Maehara CK, Chaparro JD, Lu S, Liu R, Graham A, Berry E, Hsu C-N, Kanegaye JT, Lloyd DD, Ohno-Machado L, Burns JC, Tremoulet AH, Kuehl D, editor Pediatric Emergency Medicine Kawasaki Disease Research Group. [cited 2017 Aug 16] Building a Natural Language Processing Tool to Identify Patients with High Clinical Suspicion for Kawasaki Disease from Emergency Department Notes; *Acad Emerg Med* [Internet]. 2016. May, 628–36. Available from: <http://doi.wiley.com/10.1111/acem.12925>.
 37. Murray T, Berberian L. The importance of structured data elements in EHRs. *Computer World*.
 38. Adler-Milstein J, Embi PJ, Middleton B, Sarkar IN, Smith J. Crossing the health IT chasm: considerations and policy recommendations to overcome current challenges and enable value-based care. *J Am Med Informatics Assoc* [Internet]. 2017 Sep 1; 24(5):1036–43. [cited 2017 Dec 21] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28340128>.
 39. Kern LM, Barrón Y, Dhopeswarkar RV, Edwards A, Kaushal R. HITEC Investigators. Electronic Health Records and Ambulatory Quality of Care. *J Gen Intern Med* [Internet]. 2013 Apr 3; 28(4): 496–503. [cited 2018 Jan 3] 40. Bell LM, Grundmeier R, Localio R, Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23054927>.
 40. Zorc J, Fiks AG, Zhang X, Stephens TB, Swietlik M, Guevara JP. Electronic health record- based decision support to improve asthma care: a cluster- randomized trial. *Pediatrics* [Internet]. 2010 Apr 1; 125(4):e770–7. [cited 2018 Jan 3] Available from: <http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.2009-1385>.

Subjective		ROS		Physical Exam		Plan	
Antibiotic prophylaxis	☐	Yes	No				
Bowel/Bladder dysfunction	☐	Yes	No				
Interim UTI	☐	yes	no	febrile	nonfebrile		
Lab studies at current visit or within last 30 days							
Urinalysis	☐	no	normal	proteinuria	bacteruria	leuk esterase	nitrate
Urine Culture	☐	no	positive	negative			
Findings Since Last Visit							
Ultrasound	☐	Date		Right		Left	
☐ not performed				Grade 0	Grade 1	Grade 2	Grade 3
				Grade 4			
VCUG	☐	Date		Right		Left	
☐ not performed				Grade 0	Grade 1	Grade 2	Grade 3
				Grade 4	Grade 5		
DMSA Scan	☐	Date		Right % Function	Right Presence of Scarring	Left % Function	Left Presence of scarring
☐ not performed				<10	global	<10	global
				21-30	focal	31-40	focal
				41-50	none	41-50	none
				>50		>50	
Mag-3 Scan	☐	Date		Right % Function	Right T 1/2	Left % Function	Left T 1/2
☐ not performed				<10	<10	<10	<10
				21-30	10-20	31-40	10-20
				41-50	>20	41-50	>20
				>50		>50	
PLAN							
Further followup	☐	Yes	No				
Antibiotic prophylaxis	☐	Yes	No				
Further testing including	☐	Ultrasound	VCUG	DMSA	Mag-3	MRI	RNC
on (approximately)	☐						
Antireflux Surgery	☐	Deflux	Intravesical Reimplant	Extravesical Reimplant	Minimally Invasive Reimplant		

Figure 1.

Table 1**Methodologies for Use of EHR-Based Data**

	Enhanced Lists	Direct Extraction from Pre- Existing Fields	Structured Data Elements	Unstructured Data Extraction (Natural Language Processing)
Approach	Use of ICD-9/10 codes, encounter time periods, or other standardized data to (1) generate a potential cohort of patients, (2) extract data from those data fields, and (3) abstract other relevant data using traditional chart review or other EHR data extraction methods	Extraction of data from pre-existing standard data elements, such as demographics, orders, machine-generated lab results, vital signs, appointments, and billing data	Development of customized structured data elements to allow for capture of specialty-specific data metrics	Use of sophisticated computer algorithms to convert unstructured text into structured data elements
Strengths	<p>Increased efficiency of identifying eligible patients Allows more flexibility in data extraction by coupling with traditional chart review than allowed by extraction from structured data elements</p> <p>Less upfront cost due to lack of need of specialized data element development or data extraction algorithms</p>	<p>Allows for data extraction using standardized reporting tools Flexibility of use across multiple clinic workflows Little cost due to lack of need of specialized data element development or data extraction algorithms</p> <p>May facilitate multi-center research due to lack of need for development of customized structured data elements</p>	Allows for comprehensive data capture tailored to specific research questions	Minimal impact on clinical workflow because allows for variations in clinical documentation
Weaknesses	Requires additional resources such as trained chart reviewers, structured data elements, or natural language processing for data extraction	Limited scope of data available Variations in data capture from different workflows or different sites Limited ability to obtain data from incomplete encounters or outside records Limited ability to identify temporal relationships	Upfront cost of development of customized data elements and extraction Impact on existing clinical workflow due to changes in clinical documentation Potential difficulty of expansion across multiple centers due to differences in EHR support and expertise	<p>Significant cost and expertise needed to extract meaningful data due to variations in text</p> <p>May not capture text due to uncommon abbreviations or “cut and paste” Impacted by minor changes in documentation</p> <p>Limited ability to identify temporal relationships May not easily be expanded to other centers</p>