



DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Dynamic Clinical Data Mining: Search Engine-Based Decision Support

The Harvard community has made this article openly available. [Please share](#) how this access benefits you. Your story matters.

Citation	Celi, Leo Anthony, Andrew J Zimolzak, and David J Stone. 2014. "Dynamic Clinical Data Mining: Search Engine-Based Decision Support." JMIR Medical Informatics 2 (1): e13. doi:10.2196/medinform.3110. http://dx.doi.org/10.2196/medinform.3110 .
Published Version	doi:10.2196/medinform.3110
Accessed	February 17, 2015 11:15:16 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:13890689
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

Viewpoint

Dynamic Clinical Data Mining: Search Engine-Based Decision Support

Leo Anthony Celi^{1,2}, MD, MSc, MPH; Andrew J Zimolzak³, MD, MMSc; David J Stone⁴, MD

¹Harvard-MIT Division of Health Science and Technology, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, United States

²Beth Israel Deaconess Medical Center, Boston, MA, United States

³Children's Hospital Informatics Program, Department of Pediatrics, Harvard Medical School, Boston, MA, United States

⁴University of Virginia School of Medicine, Departments of Anesthesiology and Neurosurgery, Charlottesville, VA, United States

Corresponding Author:

Leo Anthony Celi, MD, MSc, MPH

Harvard-MIT Division of Health Science and Technology

Institute for Medical Engineering and Science

Massachusetts Institute of Technology

77 Massachusetts Avenue

E25-505

Cambridge, MA, 02139

United States

Phone: 1 617 253 7937

Fax: 1 617 258 7859

Email: lceli@mit.edu

Abstract

The research world is undergoing a transformation into one in which data, on massive levels, is freely shared. In the clinical world, the capture of data on a consistent basis has only recently begun. We propose an operational vision for a digitally based care system that incorporates data-based clinical decision making. The system would aggregate individual patient electronic medical data in the course of care; query a universal, de-identified clinical database using modified search engine technology in real time; identify prior cases of sufficient similarity as to be instructive to the case at hand; and populate the individual patient's electronic medical record with pertinent decision support material such as suggested interventions and prognosis, based on prior outcomes. Every individual's course, including subsequent outcomes, would then further populate the population database to create a feedback loop to benefit the care of future patients.

(*JMIR Med Inform* 2014;2(1):e13) doi:[10.2196/medinform.3110](https://doi.org/10.2196/medinform.3110)

KEYWORDS

decision support; clinical informatics; big data

Introduction

With the near universal implementation of electronic medical records (EMRs) in conjunction with enhanced data storage options, the time nears for real time data utilization in the clinical care process [1]. The subject of the increasing importance of data for health care is much talked and written about, but there is much less discussion regarding how data might specifically be used to drive and improve the individual clinician-patient interactions that accrue to formulate the process of health care. In other words, how could complete clinical decision support be implemented across the entire health care system? Big data is an increasing presence in health care, but

data of all sizes are still underutilized. In those instances when they are used at all, they are used mainly in a retrospective analytic manner to analyze outcomes, processes, and costs. Currently, they do not dynamically drive clinical decision making in real time.

We have written on the need for the better use of intensive care unit data, noting that the development of data-based clinical decision support (CDS) tools would be one of the benefits of more comprehensive data capture [2]. Currently, the medical digital world comprises systems that are technically networked, but with data that are not systematically gathered, captured, or analyzed [3]. There are several studies that have demonstrated the potential applications and potential of capturing and

analyzing clinical data [4,5]. In a more general response to this challenge, we describe a solution that combines the utilization of three fundamental components in real time: (1) big data, (2) search engines, and (3) EMRs. In particular, search engines are brilliant tools that we all utilize many times each day; however, they have not been systematically employed for the purpose of CDS, and they represent an overlooked resource.

Dynamic Clinical Data Mining

The struggle to implement EMRs is finally coming to a mainly successful end in North America. However, the current generation of EMRs serves to digitalize information, but not to leverage it. The next step in the clinical digitization process should be the creation of a medical Internet equivalent that incorporates the rapid, powerful data search engine features that all current Web users employ. We refer to the real time incorporation of external data into the workflow as dynamic clinical data mining (DCDM) (Figure 1 shows this mining). This process will drive the design of the next generation of EMRs, and it will subsequently support the next required stage of the digital transformation process by turning medical practice into a data driven, logical, and optimized system. This care support system will provide users with the timely information that they require to make the increasingly complex decisions of medical practice.

We propose a system in which the knowledge gained from the care of individuals systematically contributes to the care of populations. The loop is closed when the richer data available in the population datasets is subsequently used in the care of individuals. DCDM would leverage the automatic crowd sourcing available in the form of population outcome analysis to formulate individualized diagnostic and therapeutic recommendations in real time. In other words, our viewpoint aligns with the Committee on a Framework for Developing a New Taxonomy of Disease, who advocate that “researchers and health care providers have access to very large sets of health and disease-related data linked to individual patients” in order to facilitate precision medicine [6]. To the Committee’s position, we would add our own that researchers and clinicians are already experienced with Internet search engines, so they would be comfortable with the identification of pertinent clinical information by accessing these large sets of data through a search engine metaphor. Currently, most clinical guidelines are generated by expert opinion based on experience and research findings such as randomized controlled trials [7]; DCDM would formulate the functional equivalent of personalized clinical guidelines.

While leading a team in the intensive care unit (ICU), one of us (LAC) experienced a difficult decision involving the resumption of anticoagulation in a patient with two mechanical heart valves. The patient was recovering from endocarditis complicated by brain abscesses. The team consulted local experts as well as the literature to guide them in weighing the risks and benefits of reinitiating anticoagulation, given the patient’s age, comorbidities, the specific bacteria involved, the number of mechanical valves, the extent and current status of the infection, etc. The information resources that were accessed

provided only general recommendations that were obviously not tailored to the patient’s demographics and comorbidities, nor to the specifics of the clinical context. The majority of these recommendations were based on expert opinions or small clinical trials, and not on “gold-standard”, multi-center randomized controlled studies. The decision was made to restart anticoagulation cautiously, given the patient’s clinical stability, the absence of bleeding complications during the acute phase, and the lack of any planned surgical intervention. In fact, preparations were underway for discharge to a skilled nursing facility. Unfortunately, four days after reinitiation of anticoagulation, the patient suffered from a massive hemorrhage of one of the brain abscesses, prompting emergent hemicraniectomy. A DCDM system could have provided predictions of the harms and benefits of anticoagulation for such a complicated patient, and it would provide the previous outcomes associated with each treatment option to review in real time [8].

Uncertainties are not limited to complex scenarios, but occur with alarming frequency in all medical settings. For example, on a daily basis in the ICU, emergency department, or the operating room, clinicians target a desired blood pressure according to population-based guidelines. When hypotension ensues, the timing, mode, and extent of intervention to maintain that goal remain art rather than science. Given that interventions to raise blood pressure such as vasopressor therapy or fluids are associated with risk of harm if given even slightly in excess, it is crucial that the targeted blood pressure be personalized as much as possible. DCDM would add the knowledge gained from prior care of populations to the current local data specifics in order to formulate an approach that is optimal in terms of both the short-term goal as well as the long-term outcomes. For instance, DCDM could assist an ICU physician in choosing an intervention and its dose to treat shock, such that the intervention has the optimal effect on the short-term blood pressure profile and long-term mortality, length of stay, and/or eventual quality of life.

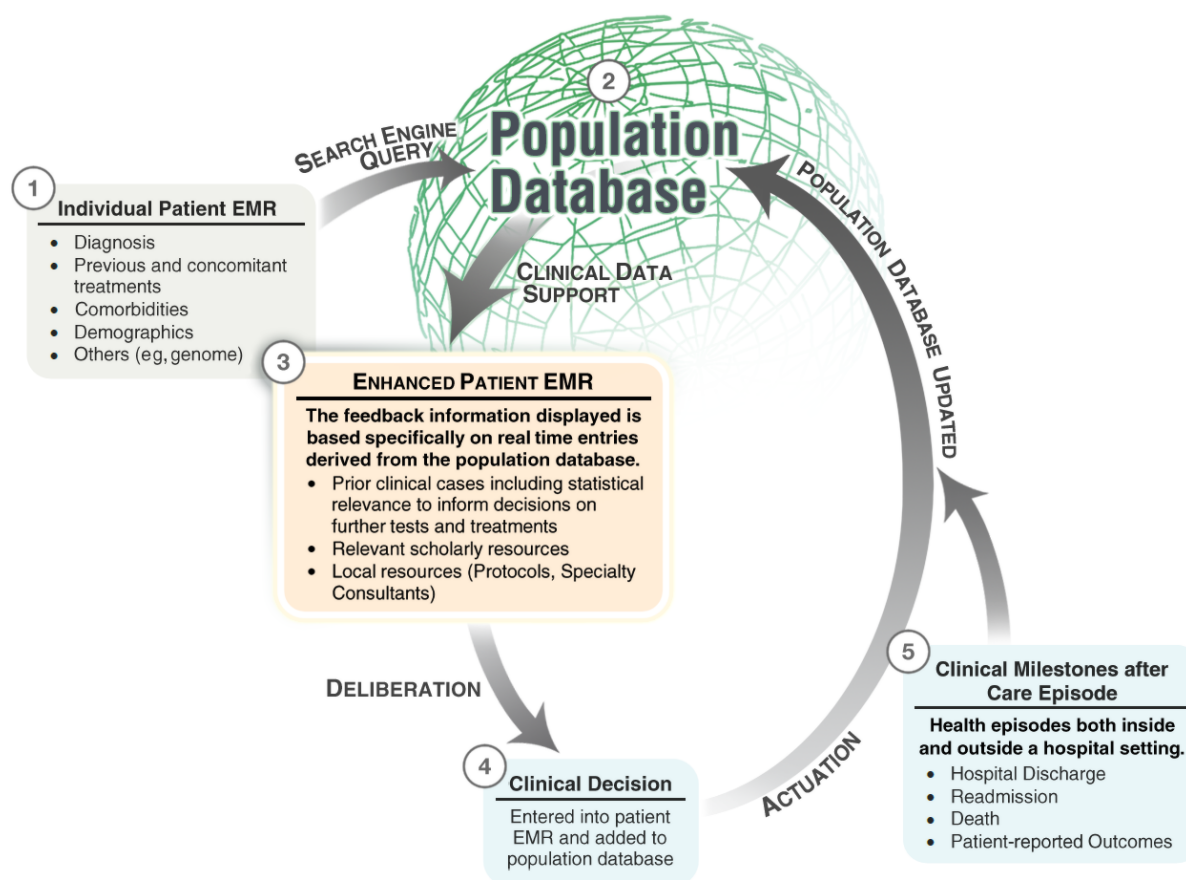
Other studies have explored similar themes. Certainly the application of logic and probability to medicine has been discussed for decades [9]. More recently, and more to the point of our discussion, a variety of commentators have called for a nationwide learning health system [10,11]. In 2011, Frankovich et al reported the case of a girl with lupus and potential thrombotic risk factors. To determine whether anticoagulation was appropriate, they used text searching to retrieve records of similar patients from their hospital’s EMR, followed by a focused manual review. They found that pediatric patients with lupus and these potential risk factors indeed had a higher risk of thrombosis than those without the risk factors, and they elected to start anticoagulation as a result [12]. The Query Health initiative, from the Office of the National Coordinator for Health Information Technology, intends to facilitate distributed queries, which can aggregate results from multiple organizations’ patient populations while preserving data security [13]. Similarly, the goal of the Strategic Health Information Technology Advanced Research Project on secondary use (SHARPN) is to standardize structured and unstructured EMR data to promote its reuse [14]. The open source Clinical Text Analysis and Knowledge

Extraction System and the SHARPN program use the Unstructured Information Management Architecture, the same architecture that allowed the Watson system from International Business Machines to compete on the Jeopardy! television quiz show [15,16]. In general, search engines for unstructured text are seen as the first step, and implementation of “content analytics” is the next step to extract information, allow exploration, and to improve search [17].

CDS provides caregivers with information to improve the quality of their decision making, yet caregivers still do not have available a dynamic, comparative analysis between the current patient and all available data generated during clinical care. This analysis is individually tailored because it uses EMR data

entered on one specific patient, yet it remains population-based because the analysis makes a comparison to population data, to identify similar clinical situations from the past, and to mine them for interventions and subsequent outcomes (as illustrated in point 5 in Figure 1). Thus, the clinician does not have to make a decision in isolation from what has been tried, observed, and documented by many colleagues in many other similar patients. In addition, the information provided would provide useful support to the process of patient-physician shared decision making [18]. This approach would interrogate data to suggest next step options and weigh the risks and benefits of a treatment or test for a specific patient, the Holy Grail of personalized medicine.

Figure 1. Dynamic clinical data mining. Figure courtesy of Kai-ou Tang and Edward Moseley. EMR=electronic medical record.



Discussion

Required Data and Information

The most basic requirement for the DCDM system is the complete digital capture of patient information. We would maintain that de-identified clinical data constitutes a public good and should reside in a carefully managed public domain database, overseen by a cooperative coalition of vendors, provider institutions, and regulators. This is already the case for federally funded research data in the United States, and a movement is underway to share participant-level data from clinical trials [19-21]. Furthermore, the Patient Centered Outcomes Research Institute in the United States has already

begun to develop the infrastructure that will aggregate large amounts of de-identified patient data from diverse sources for the purposes of observational research studies [22]. Any central database or federated query system must of course be governed by policies that account for the interests and preferences of the public regarding patient privacy, and the purposes for which the data are used [23]. The costs of database management would be built into purchase and maintenance agreements. Subsequent analyses would identify the clinical and financial impact of the entire data-based system with adjustments made as necessary.

In a DCDM system, a search engine would accept both structured and unstructured search terms to query the population database, much as current search engines query the database of

the entire Internet. The unstructured terms could be used in a query via real time natural language processing or the next generation of “text to code” conversion applications, which convert free text to coded, structured search terms, while considering the context provided by free text, in order to ensure accuracy and clinical intention. The individual’s data would be rapidly compared to the population database to capture a set of useful records that match the content and context of the care encounter.

We envision every patient’s health data digitally catalogued according to demographics, diagnoses, treatments, and outcomes, all time stamped for sequential interpretation. We suspect the types of data included will evolve rapidly over time. For example, future data may be derived from cell phones or home monitors. This will be the basis of a data-based learning system of care where choices are made on the basis of substantial data, statistical support programs, and documented outcomes, rather than on individual experience and inconsistent use of applicable informational resources.

Potential Obstacles

A significant caveat is that bias and/or residual confounding by indication may mar the analysis. The goal is to identify patient records in the database that are as similar as possible to the patient in terms of the variables that can confound the relationship between the intervention and the outcome as identified by clinician heuristics and complemented by computer algorithms, and then to compare the outcomes of those who receive the intervention versus those who did not. Residual confounding means that the outcome difference might not be due to the intervention, but rather due to something inherent to those patients who receive the treatment, or their condition. Realizing that the system is to be used by clinicians rather than data scientists, it must be designed so that such confounding and bias are minimized, with the confidence levels around the estimate of the treatment effects quantified and explained at the clinical user interface level.

The use of raw data from a variety of sources will present challenges. We acknowledge the inherent heterogeneity of people and disease. This presents an issue in terms of the levels of detail that require capture. The integration of data from multiple sources will require the use of standard terminologies and ontologies to allow for compatibility of the data from one source to another [14]. With the use of such standardization, these heterogeneities become inconveniences, not obstacles, to the vision. We foresee the implementation of progressively better EMRs, networks, and databases, all used by a generation of clinicians who have grown up with, are comfortable with, and expect to use and benefit from digital tools. It is important to anticipate potential risks, but this should be done in order to design and build the system so as to minimize them.

CDS tools must be engineered purposefully into workflow to avoid actually increasing user time and work requirements. An author of this paper (LAC) has previously reported on the use of local databases for the creation of CDS tools [24-26], which is one of the “grand challenges” in CDS [27]. Recent work adds the input of dynamic variables, which capture more information than traditional prediction models, including data on changes

and variability of repeatedly measured values [28]. The readers are hereby directed to a recent review of the use of data mining in CDS [29]. DCDM would extend the capabilities of CDS by dynamically incorporating both individual and population data in real time [30,31]. In addition to querying and populating local databases, DCDM would also use the power of search engine technology to leverage population level data.

Organization and Actuation

Combined clinical and engineering teams would need to work together to generate algorithms to determine the weight of each feature being matched against the outcome of interest, as well as the relative value of (and permissible missing values for) the interacting data elements in the match process. These algorithms should be modifiable in order to meet the continuously changing practice of medicine. Search engine algorithms are modifiable and these modifications can be engineered for specific purposes. Google has made a number of such strategic modifications to its algorithms over time [32]. It is likely that a prototype employing a smaller search target such as the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) Database [33] would be required to demonstrate the practicality and utility of the concept, as well as to create, develop, and initially refine the search engine algorithm. Indeed, the MIMIC Database has been previously employed to predict fluid responsiveness among hypotensive patients [34], as well as the hematocrit trend among patients with gastrointestinal bleeding [35], using the trajectory of physiologic variables over time.

The system would identify and suggest prioritized interventions and other courses of action that have been shown to be most valuable in terms of outcome and cost. The system’s features might include displays of quantitative and qualitative description of the match, hyperlinks that allow the user to drill further into the underlying data that is returned, and links to conventional practice guidelines and evidence-based modalities.

A clinical decision must be made at one point in time, but in most cases, decisions are ongoing and iterative. The system will incorporate the short-term response to the prior intervention each time that the system is accessed, but also capture long-term outcomes. For example, when a physician orders an intervention in response to acute kidney injury, the system would log the short-term response in serum creatinine, and also the long-term outcome of progression to or prevention of end-stage renal disease. The system could also be independently data mined to identify patterns that indicate whether the patient course is on track toward the desired outcome.

Large, diverse, international populations would improve the opportunity to achieve matches. When no match is possible, an alert could be provided noting the unusual features that preclude a match. The system would then provide appropriate suggestions for the user, such as a specialty referral, a data error of some kind, or even the possible detection of an entirely new condition. It would also serve as an epidemiological tool that recognizes emerging or spreading contagions [36,37], or other harmful exposures [38,39] more quickly and efficiently than is currently possible.

Conclusion

DCDM has its roots in the need for medical care to be more fully based on data. The universal collection of data would also present the additional advantages of providing future opportunities to formulate randomized registry trials, as well as for other directed data mining purposes [40]. DCDM would begin to transform the exigent data entries that clinicians

perform on a daily basis into a real tool for clinical care. Decisions would be made on the basis of experience over vast populations, rather than solely on individual knowledge and experience. We propose the creation of a system that supports clinician decision makers so that their decisions can be as logical, transparent, and unambiguous as possible. DCDM would more gainfully employ the power of networked computers, search engines, and data storage advances to leverage the copious, but underused data entered into EMRs.

Acknowledgments

We would like to thank Marie Csete, MD, PhD; and Daniel Stone, PhD, for their helpful comments.

Conflicts of Interest

None declared.

References

1. Rodrigues JJ, de la Torre I, Fernández G, López-Coronado M. Analysis of the security and privacy requirements of cloud-based electronic health records systems. *J Med Internet Res* 2013;15(8):e186 [FREE Full text] [doi: [10.2196/jmir.2494](https://doi.org/10.2196/jmir.2494)] [Medline: [23965254](https://pubmed.ncbi.nlm.nih.gov/23965254/)]
2. Celi LA, Mark RG, Stone DJ, Montgomery RA. "Big data" in the intensive care unit. Closing the data loop. *Am J Respir Crit Care Med* 2013 Jun 1;187(11):1157-1160 [FREE Full text] [doi: [10.1164/rccm.201212-2311ED](https://doi.org/10.1164/rccm.201212-2311ED)] [Medline: [23725609](https://pubmed.ncbi.nlm.nih.gov/23725609/)]
3. The Committee on the Learning Health Care System in America. In: Smith M, Saunders R, Stuckhardt L, McGinnis JM, editors. Best care at lower cost: The path to continuously learning health care in America. USA: The National Academies Press; 2013.
4. Fine AM, Nigrovic LE, Reis BY, Cook EF, Mandl KD. Linking surveillance to action: Incorporation of real-time regional data into a medical decision rule. *J Am Med Inform Assoc* 2007;14(2):206-211 [FREE Full text] [doi: [10.1197/jamia.M2253](https://doi.org/10.1197/jamia.M2253)] [Medline: [17213492](https://pubmed.ncbi.nlm.nih.gov/17213492/)]
5. Davis DA, Chawla NV, Christakis NA, Barabási A. Time to CARE: A collaborative engine for practical disease prediction. *Data Min Knowl Disc* 2009 Nov 25;20(3):388-415. [doi: [10.1007/s10618-009-0156-z](https://doi.org/10.1007/s10618-009-0156-z)]
6. Committee on a Framework for Development a New Taxonomy of Disease, Board on Life Sciences, Division on Earth and Life Studies, National Research Council. Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease. USA: National Academies Press; 2011.
7. Carville S, Harker M, Henderson R, Gray H, Guideline Development Group. Acute management of myocardial infarction with ST-segment elevation: Summary of NICE guidance. *BMJ* 2013;347:f4006. [Medline: [23843545](https://pubmed.ncbi.nlm.nih.gov/23843545/)]
8. Bennett C, Doub T. caseybennett.com. 2010. Data mining and electronic health records: Selecting optimal clinical treatments in practice URL: http://www.caseybennett.com/uploads/DMIN_Paper_Proposal_Final_Arxiv.pdf [accessed 2014-06-14] [WebCite Cache ID 6QKOCNab]
9. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science* 1959 Jul 3;130(3366):9-21. [Medline: [13668531](https://pubmed.ncbi.nlm.nih.gov/13668531/)]
10. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med* 2010 Nov 10;2(57):57cm29. [doi: [10.1126/scitranslmed.3001456](https://doi.org/10.1126/scitranslmed.3001456)] [Medline: [21068440](https://pubmed.ncbi.nlm.nih.gov/21068440/)]
11. Roundtable on Value & Science-Driven Health Care, The Learning Health System Series, Institute of Medicine. In: Grossman C, McGinnis JM, editors. Digital infrastructure for the learning health system: The foundation for continuous improvement in health and health care: Workshop series summary (the learning health system series). USA: National Academies Press; 2011.
12. Frankovich J, Longhurst CA, Sutherland SM. Evidence-based medicine in the EMR era. *N Engl J Med* 2011 Nov 10;365(19):1758-1759. [doi: [10.1056/NEJMp1108726](https://doi.org/10.1056/NEJMp1108726)] [Medline: [22047518](https://pubmed.ncbi.nlm.nih.gov/22047518/)]
13. Klann JG, Buck MD, Brown J, Hadley M, Elmore R, Weber GM, et al. Query health: Standards-based, cross-platform population health surveillance. *J Am Med Inform Assoc* 2014 Jul;21(4):650-656. [doi: [10.1136/amiajnl-2014-002707](https://doi.org/10.1136/amiajnl-2014-002707)] [Medline: [24699371](https://pubmed.ncbi.nlm.nih.gov/24699371/)]
14. Pathak J, Bailey KR, Beebe CE, Bethard S, Carrell DC, Chen PJ, et al. Normalization and standardization of electronic health records for high-throughput phenotyping: The SHARPN consortium. *J Am Med Inform Assoc* 2013 Dec;20(e2):e341-e348. [doi: [10.1136/amiajnl-2013-001939](https://doi.org/10.1136/amiajnl-2013-001939)] [Medline: [24190931](https://pubmed.ncbi.nlm.nih.gov/24190931/)]
15. Chute CG, Pathak J, Savova GK, Bailey KR, Schor MI, Hart LA, et al. The SHARPN project on secondary use of electronic medical record data: Progress, plans, and possibilities. *AMIA Annu Symp Proc* 2011;2011:248-256 [FREE Full text] [Medline: [22195076](https://pubmed.ncbi.nlm.nih.gov/22195076/)]

16. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010;17(5):507-513 [FREE Full text] [doi: [10.1136/jamia.2009.001560](https://doi.org/10.1136/jamia.2009.001560)] [Medline: [20819853](https://pubmed.ncbi.nlm.nih.gov/20819853/)]
17. Feldman S, Reynolds H, Schubmehl D. Content analytics and the high-performing enterprise. Framingham, MA: IDC; 2012. URL: http://public.dhe.ibm.com/software/data/sw-library/ecm-programs/ContentAnalytics_IDC.pdf [accessed 2014-06-05] [WebCite Cache ID [6Q6xdxLqK](https://www.webcitation.org/6Q6xdxLqK)]
18. Oshima Lee E, Emanuel EJ. Shared decision making to improve care and reduce costs. *N Engl J Med* 2013 Jan 3;368(1):6-8. [doi: [10.1056/NEJMp1209500](https://doi.org/10.1056/NEJMp1209500)] [Medline: [23281971](https://pubmed.ncbi.nlm.nih.gov/23281971/)]
19. Holdren JP. Executive Office of the President, Office of Science and Technology Policy. Increasing access to the results of federally funded scientific research Internet URL: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf [accessed 2013-11-18] [WebCite Cache ID [6LEyusHy6](https://www.webcitation.org/6LEyusHy6)]
20. Mello MM, Francer JK, Wilenzick M, Teden P, Bierer BE, Barnes M. Preparing for responsible sharing of clinical trial data. *N Engl J Med* 2013 Oct 24;369(17):1651-1658. [doi: [10.1056/NEJMp1309073](https://doi.org/10.1056/NEJMp1309073)] [Medline: [24144394](https://pubmed.ncbi.nlm.nih.gov/24144394/)]
21. Drazen JM. Open data. *N Engl J Med* 2014 Feb 13;370(7):662. [doi: [10.1056/NEJMe1400850](https://doi.org/10.1056/NEJMe1400850)] [Medline: [24450856](https://pubmed.ncbi.nlm.nih.gov/24450856/)]
22. Selby J. Patient-centered outcomes research institute. Building PCORnet as a secure platform for outcomes research URL: <http://www.pcori.org/blog/building-pcor-net-as-a-secure-platform-for-outcomes-research/> [accessed 2014-05-13] [WebCite Cache ID [6PXCikL4I](https://www.webcitation.org/6PXCikL4I)]
23. Grande D, Mitra N, Shah A, Wan F, Asch DA. Public preferences about secondary uses of electronic health information. *JAMA Intern Med* 2013 Oct 28;173(19):1798-1806. [doi: [10.1001/jamainternmed.2013.9166](https://doi.org/10.1001/jamainternmed.2013.9166)] [Medline: [23958803](https://pubmed.ncbi.nlm.nih.gov/23958803/)]
24. Lee J, Kothari R, Ladapo JA, Scott DJ, Celi LA. Interrogating a clinical database to study treatment of hypotension in the critically ill. *BMJ Open* 2012;2(3) [FREE Full text] [doi: [10.1136/bmjopen-2012-000916](https://doi.org/10.1136/bmjopen-2012-000916)] [Medline: [22685222](https://pubmed.ncbi.nlm.nih.gov/22685222/)]
25. Celi LA, Tang RJ, Villarroel MC, Davidzon GA, Lester WT, Chueh HC. A clinical database-driven approach to decision support: Predicting mortality among patients with acute kidney injury. *J Healthc Eng* 2011 Mar;2(1):97-110 [FREE Full text] [doi: [10.1260/2040-2295.2.1.97](https://doi.org/10.1260/2040-2295.2.1.97)] [Medline: [22844575](https://pubmed.ncbi.nlm.nih.gov/22844575/)]
26. Celi LA, Galvin S, Davidzon G, Lee J, Scott D, Mark R. A database-driven decision support system: Customized mortality prediction. *J Pers Med* 2012 Sep 27;2(4):138-148 [FREE Full text] [doi: [10.3390/jpm2040138](https://doi.org/10.3390/jpm2040138)] [Medline: [23766893](https://pubmed.ncbi.nlm.nih.gov/23766893/)]
27. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. *J Biomed Inform* 2008 Apr;41(2):387-392 [FREE Full text] [doi: [10.1016/j.jbi.2007.09.003](https://doi.org/10.1016/j.jbi.2007.09.003)] [Medline: [18029232](https://pubmed.ncbi.nlm.nih.gov/18029232/)]
28. Mayaud L, Lai PS, Clifford GD, Tarassenko L, Celi LA, Annane D. Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension. *Crit Care Med* 2013 Apr;41(4):954-962 [FREE Full text] [doi: [10.1097/CCM.0b013e3182772adb](https://doi.org/10.1097/CCM.0b013e3182772adb)] [Medline: [23385106](https://pubmed.ncbi.nlm.nih.gov/23385106/)]
29. Aleksovska-Stojkowska L, Loskovska S. Data mining in clinical decision support systems. In: Gaol FL, editor. Recent progress in data engineering and internet technology. Berlin: Springer; 2013:287-293.
30. Friedman CP, Rubin JC. The learning system: Now more than ever - Let's achieve it together!. 2014 Feb 25 Presented at: HIMSS 14 Annual Conference & Exhibition; 2014-02-25; Orlando, Florida URL: <http://himss.files.cms-plus.com/2014Conference/handouts/70.pdf> [WebCite Cache]
31. Clinical personalized pragmatic predictions of outcomes. Department of Veterans Affairs URL: <http://www.clinical3po.org/> [accessed 2014-05-13] [WebCite Cache ID [6PXBjQHPN](https://www.webcitation.org/6PXBjQHPN)]
32. Sullivan D. Search Engine Land. FAQ: All about the new Google "hummingbird" algorithm URL: <http://searchengineland.com/google-hummingbird-172816> [accessed 2013-11-18] [WebCite Cache ID [6LEYRMjZ](https://www.webcitation.org/6LEYRMjZ)]
33. PhysioNet. 2012. MIMIC II Databases URL: <http://physionet.org/mimic2/> [accessed 2013-11-18] [WebCite Cache ID [6LEYagQz1](https://www.webcitation.org/6LEYagQz1)]
34. Fialho AS, Celi LA, Cismondi F, Vieira SM, Reti SR, Sousa JM, et al. Disease-based modeling to predict fluid response in intensive care units. *Methods Inf Med* 2013;52(6):494-502. [doi: [10.3414/ME12-01-0093](https://doi.org/10.3414/ME12-01-0093)] [Medline: [23986268](https://pubmed.ncbi.nlm.nih.gov/23986268/)]
35. Cismondi F, Celi LA, Fialho AS, Vieira SM, Reti SR, Sousa JM, et al. Reducing unnecessary lab testing in the ICU with artificial intelligence. *Int J Med Inform* 2013 May;82(5):345-358. [doi: [10.1016/j.ijmedinf.2012.11.017](https://doi.org/10.1016/j.ijmedinf.2012.11.017)] [Medline: [23273628](https://pubmed.ncbi.nlm.nih.gov/23273628/)]
36. Gesteland PH, Gardner RM, Tsui FC, Espino JU, Rolfs RT, James BC, et al. Automated syndromic surveillance for the 2002 Winter Olympics. *J Am Med Inform Assoc* 2003;10(6):547-554 [FREE Full text] [doi: [10.1197/jamia.M1352](https://doi.org/10.1197/jamia.M1352)] [Medline: [12925547](https://pubmed.ncbi.nlm.nih.gov/12925547/)]
37. Elkin PL, Froehling DA, Wahner-Roedler DL, Brown SH, Bailey KR. Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes. *Ann Intern Med* 2012 Jan 3;156(1 Pt 1):11-18. [doi: [10.7326/0003-4819-156-1-201201030-00003](https://doi.org/10.7326/0003-4819-156-1-201201030-00003)] [Medline: [22213490](https://pubmed.ncbi.nlm.nih.gov/22213490/)]
38. Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G. Detecting adverse events using information technology. *J Am Med Inform Assoc* 2003;10(2):115-128 [FREE Full text] [Medline: [12595401](https://pubmed.ncbi.nlm.nih.gov/12595401/)]
39. Brownstein JS, Sordo M, Kohane IS, Mandl KD. The tell-tale heart: Population-based surveillance reveals an association of rofecoxib and celecoxib with myocardial infarction. *PLoS One* 2007;2(9):e840 [FREE Full text] [doi: [10.1371/journal.pone.0000840](https://doi.org/10.1371/journal.pone.0000840)] [Medline: [17786211](https://pubmed.ncbi.nlm.nih.gov/17786211/)]
40. Lauer MS, D'Agostino RB. The randomized registry trial--the next disruptive technology in clinical research? *N Engl J Med* 2013 Oct 24;369(17):1579-1581. [doi: [10.1056/NEJMp1310102](https://doi.org/10.1056/NEJMp1310102)] [Medline: [23991657](https://pubmed.ncbi.nlm.nih.gov/23991657/)]

Abbreviations

CDS: clinical decision support

DCDM: dynamic clinical data mining

EMRs: electronic medical records

ICU: intensive care unit

MIMIC: Multiparameter Intelligent Monitoring in Intensive Care

SHARPN: Strategic Health Information Technology Advanced Research Project on secondary use

Edited by G Eysenbach; submitted 19.11.13; peer-reviewed by D Maslove, S Seevanayanagam, J Klann, A Holzinger, WC Su, J Turner; comments to author 27.01.14; revised version received 28.02.14; accepted 22.05.14; published 23.06.14

Please cite as:

Celi LA, Zimolzak AJ, Stone DJ

Dynamic Clinical Data Mining: Search Engine-Based Decision Support

JMIR Med Inform 2014;2(1):e13

URL: <http://medinform.jmir.org/2014/1/e13/>

doi: [10.2196/medinform.3110](https://doi.org/10.2196/medinform.3110)

PMID:

©Leo Anthony Celi, Andrew J Zimolzak, David J Stone. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org>), 23.06.2014. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.