



Published in final edited form as:

Curr Opin Crit Care. 2014 October ; 20(5): 573–580. doi:10.1097/MCC.0000000000000137.

Optimal Data Systems: The future of clinical predictions and decision support

Leo Anthony Celi, Marie Csete, and David Stone

Abstract

Purpose of Review—The purpose of the review is to describe the evolving concept and role of data as it relates to clinical predictions and decision-making.

Recent Findings—Critical care medicine is, as an especially data rich specialty, becoming acutely cognizant both of its historic deficits in data utilization but also of its enormous potential for capturing, mining, and leveraging such data into well designed decision support modalities as well as the formulation of robust best practices.

Summary—Modern electronic medical records create an opportunity to design complete and functional data systems that can support clinical care to a degree never seen before. Such systems are often referred to as ‘data-driven’ but a better term is ‘optimal data systems’. Here we discuss basic features of an optimal data system and its benefits, including the potential to transform clinical prediction and decision support.

Keywords

electronic health records; data mining; decision support systems, clinical; information systems

I. Introduction: Systems of Data

The ‘age of information’ combined with ubiquitous electronic medical records (EMRs) means, in theory, that all data necessary for optimal diagnosis, treatment and prognostication can be available to clinicians. The EMR interfaced to scientific information creates both opportunity and considerable challenges in acquisition and presentation of the clinically-relevant data, in ways that best inform decision-making. Within a single patient EMR, myriad data types are captured, identified and categorized, filtered, summarized and then employed to construct a dynamic and revisable assessment and treatment plan. The amount of data generated by a single patient in a single hospital admission, particularly in the intensive care unit (ICU), is enormous. Currently, the way vast data are captured and entered into medical records, leveraged, and fed back to clinicians is far from optimal. Despite application of computational tools to support decision making in similarly data-rich complex systems outside medicine, application of computational tools to clinical data is in its infancy. Care must be taken to design such systems strategically, with sufficient modifiability to accommodate innovative advances as novel data elements and underlying decisional

Corresponding Author: Leo Anthony Celi, Massachusetts Institute of Technology, 77 Massachusetts Avenue, E25-505, Cambridge, MA 02139, USA, +1(617) 253-7937, lceli@mit.edu.

The following authors contributed equally to this manuscript: Leo Anthony Celi, Marie Csete and David Stone.

principles are added, changed and deleted from the canon. The organization of clinical data systems, then, requires a framework architecture on which data at all levels of resolution can be logically arranged. Highly functional complex systems (both engineered and evolved) share common design features that should be considered in the rational design of clinical data systems. Such meticulously-designed systems will usher in a new era in clinical predictions: the interest will expand from predicting outcomes at the patient level either for prognostication or to inform decisions, to predicting information gain from diagnostic tests and response to various treatment options for individual patients.

Arguably data from outside the EMR can and should inform clinical decision-making. For example, continuous local pollution levels play a role in health, but it is impractical to feed these back to physicians because no one knows what to do with them. In the future, one could envision a huge amount of information used for clinical purposes--various 'omics' databases, large longitudinal epidemiologic studies, clinical trials, basic and preclinical research--all (automatically) interfaced with the EMR and exploited for minute-to-minute predictions and decision support. But such a goal remains distant for now, and physicians continue to use clinical data items much the way they did during the unconnected, paper-based world of the 20th century.

Here it is critical to note that, though the amount of data (collected but) hidden from clinical records is problematic, more data does not necessarily yield better predictions, decisions or outcomes. Data organization around design principles is the key. For example, a list of every component of an airplane does not automatically yield a robust flying machine without engineering principles and controls. Similarly, future clinical record systems must be engineered with standardization at the core, customizability at the edges, the ability to accommodate changes in healthcare environment, and a software architecture that is robust and current yet modifiable without undue difficulties. Thus, while an optimized data-based care system is an ideal goal, its benefits are limited by the data available to the system, but more importantly by how the data are organized. For this reason, we focus on some near-term approaches to restructuring clinical data, as a system, from content that is currently available but not optimally employed in the context of decision-making.

Clinical care is based on data acquisition and analysis, but is not yet 'data-driven' in the stricter sense of being objective, systematic, structured and replicable with the same best outcomes. In fact, the data deluge of clinical practice (and the medical literature) has made it progressively more difficult to be aware of all applicable data. Unpredictable outcomes - specifically those relating to interventions providing no value added to the patient, or worse, consequences - are far too common and do not lend themselves easily to medicine as an applied data science. Here we describe the current state of clinical data in an attempt to clarify and enhance the concept of what has often been referred to as a *data-driven care system* in order to leverage computing power to cope with, manage, and properly analyze just the right patient data in the context of the population. The ultimate goal is to improve decision-making for physicians and patients by providing predictions and individualized recommendations in order to reliably optimize patient outcomes.

II. Architecture of a Clinical Data System

A system is an interconnected and interacting assembly of components (a.k.a. modules, parts) that can perform functions not possible with just the individual components. The rules (or protocols) that dictate the range of behaviors of a system are designed in engineered systems, and evolved in biological systems. A system accepts inputs and processes them into outputs. The details of a controlled system's sensing, computation, and actuation are dictated by the particular architecture of that system.

Clinical data tend to consist mainly of modular elements. (Note that modular elements can be descriptive or diagnostic in nature, as well as therapeutic or interventional.) Clinical data format, however, is (increasingly) highly varied (single nucleotide polymorphisms, transcriptomes of a tumor biopsy, functional imaging, raw vs. transformed EEG signals, results of a diagnostic nerve block) and therefore difficult to integrate without new collection and analytic tools. For inpatients, data are collated in a per-stay medical record along with varying degrees of accompanying interpretation. For outpatients, data are often dispersed, less well organized and often functionally unavailable. For clinical data generally, no framework architecture is used for organizing data in the context of a physiologic system (e.g. neurologic) or a medical condition (e.g. sepsis).

The advent of enterprise EMRs that incorporate outpatient and inpatient functions has begun to address the issue of integrating the patient's entire data history. Nonetheless, in every encounter with a patient, the clinician's data view axis is restricted to the prior and current data of an individual patient, as well as to the education, experience, efficiency, and memory of the clinician (or clinical team). To a large and unacceptable degree, clinicians 're-invent the wheel' with every patient encounter.

The clinician is the controller of a clinical data system, and the patient is – in engineering terms – the “plant” (Figure 1). Thinking of the clinician as controller highlights the need to structure the input data for optimal output (diagnosis, intervention, collection of more data and prediction). For the most part, the controller is the cerebral ‘wetware’ of the clinician but expert analysis will be increasingly supplemented by automated clinical decision support modalities. [1] Further sensing of the patient state (the plant) in response to actuation is fed back to the clinician. In engineered control systems (such as a thermostat) controllers are designed to iteratively re-examine and re-apply solutions to the ‘plant’, a design that is also used by the clinician with feedback from treatment response incorporated back to close the loop.

Understanding clinical data (not just clinical care in the larger sense) in terms of optimized controllers is a fundamentally important concept for clinical data utility, as (healthy) physiology is dependent on well-studied physiologic control systems. Given the gap between a well-controlled system and current clinical practice, we propose the term ‘optimal data system’ (ODS) to distinguish current data-driven approaches from those that are purposefully designed. We envision ODS as an enhanced type of data-driven system which selectively employs appropriate data elements to support formulation of the best possible decisions, including outcome prediction. These data include not only the patient's own

historical and current data, but will eventually incorporate pertinent population data findings, as well as decision support resources such as guidelines, preferably formulated without undue industrial or financial influence. [2] The idealized ODS would continuously assess and catalogue the resultant outcomes of clinical decisions to determine what are the best data and decisions that can be recommended in the future.

The ideal data system would also be organized in modules representing particular organs or disease states with these modules nested in the global data set reflecting system pathophysiology. Such organization and presentation are now in the hands of software developers, both a challenge and an opportunity. The ideal data system organization will require intense feedback between these computer scientists and clinicians for the next generation of health information systems.

III. Constraints and Tradeoffs in the Utilization of Clinical Data

Medical care systems provide caregivers with various levels of opportunity to identify and acquire the data perceived as necessary at any given time for a given patient, but data identification is highly constrained. First, the data must be *conceived* and recognized as such i.e. identified as a clinical data element. Until the element has been established to be relevant to clinical care, the element will remain in the area of the Venn diagram that lies outside the clinical data area. (Figure 2) This may seem trivial, but once-essential data can become obsolete and completely new and unexpected forms of data become essential (e.g. troponins for myocardial infarction diagnosis today versus ~1978). Data may also simply be unknown to the user because of educational, experiential or communication issues. Second, it must be recognized as *valuable* i.e. worthy of the cost of the acquisition and storage. Ideally, this value is established by studies that examine information gain of this particular element - Does it lead to a better understanding of the disease process on top of what is already known and/or does it inform decision regarding a possible intervention that will alter patient outcome? Third, the data must be *obtainable*. The data may not be technically available because of a lack of equipment or because science has not yet established a method of examining the real-time function of a given gene or signal transduction pathway. Fourth, the data must be *presented and formatted to the user in a timely manner (based on clinical acuity) and stored* for future clinical and research utilization; these functions are facilitated and supported by EMRs. Much information is simply lost because it is not archived with the patient record (e.g. waveform signals, hemodialysis parameters) or functionally inaccessible in mounds of paper or microfilm.

Currently, the predominance of free text entry in physician notes makes the reliable cataloguing of data for future analysis for prediction and decision support rather difficult, but at least theoretically possible via tools such as natural language processing. The seamless integration of structured data capture in EMR workflow is still in early development. Medical research continues to identify novel and previously unrecognized elements from the general universe of data which become relevant to clinical care. Entirely new types of data may be developed in this fashion e.g. genetic testing for disease risk.

A need for “all the data, all the time” can be wasteful, costly, confusing and time-consuming. Clinicians who require an MRI for the evaluation of all back pain or headache cases will struggle to operate in environments where these are not available (or allowed), highlighting the important issues of cost, value and risk of data. If one can function safely without such sophisticated diagnostic modalities in most circumstances, what is the appropriate threshold for using such modalities? At the other end of the spectrum, patients may suffer consequences for clinicians not performing tests. Clearly, we have not learned to capture ‘just enough data’ which should be the goal of data systems design and evaluation.

Data may also be erroneous for a variety of reasons such as mis-entry, machine or human errors, unduly subjective circumstances, and limitations of medical device precision. Finally, data may go missing either because it was never entered or was lost in some quantum mechanical event occurring in a vast database over long time periods. Clearly, we, as all too human clinicians, need some help in identifying and utilizing data optimally.

IV. Creating an Optimal Data System

Acquisition of the necessary data elements as well as their subsequent assembly both represent essential protocols of a data-driven system. The clinical puzzle is simply not as perfect as jigsaw pieces out of the box. Instead pieces are missing and misshapen, and there may be strange extra pieces. (Figure 3) Even the final puzzle product can be a moving target. However, the formulation of some kind of mental construct built on data pieces is a useful model for the next steps of assessment, intervention, and re-assessment. The ODS must be designed to support and facilitate an increasingly complex, rushed, and demanding clinical work environment.

We propose an enhancement to the current process of data incorporation into the decision and care process. Since the nexus of clinical decisions is the medical note, the EMR is the logical platform in this development. An ideal ODS would include the following:

1. Automatic collection and display of newly available data (i.e. data not yet entered in an EMR) required to complete the clinical picture. These could include patient-entered data; data sent by pre-hospital personnel real-time; and data from wearable sensors.
2. Capture and integration of the newly available and historical data along with real time physician entries (notes) to progressively characterize the clinical state and query both population database and clinical decision support modalities. These are represented by our dynamic clinical data mining (DCDM) concept and the IBM Watson™ type of functionality, respectively. [1,3] This complex feature requires de-identified data sharing on a universal basis. [3]
3. An innovative additional feature would be required to integrate the DCDM and Watson™ functions and deliver the following: Diagnostic, therapeutic, prognostic as well as further documentation suggestions would be incrementally displayed on the basis of the combination of the analyzed data provided by these multiple sources. These might include suggestions to supplement required missing data with additional testing; clarification of free text entries for purposes of standard coding;

identification of suggestive but otherwise difficult-to-identify patterns and constellations of data; automatic highlighting of diagnoses, treatments, results and combinations of results that are incongruous or inconsistent; and providing population based but individualized suggestions for ongoing care decisions and next steps. This is the stage where the software-wetware integration process is continuously enhanced by leveraging information outside the purview of today's clinical EMR (or paper chart) user.

4. Machine learning would be employed to continuously improve the quality of the information presented to the user as the system 'learns' how clinicians employ the system in heterogeneous ways.
5. Users are allowed to customize their own version of the application to the extent that standardization of data is not violated. In other words, the application design should be 'customized at the edges but standardized at the core' enabling users to have considerable but reasonable control over their interactions with the system. [4] Customization should not be permitted to the extent that it is difficult or near impossible for software engineers to investigate reported system errors and unanticipated events.
6. Saved system data would then be provided to both the local and the population databases for ongoing analysis for real time care and the objective formulation of clinical support modalities including practice guidelines and research.
7. Reports would be generated regarding user decisions in terms of consistency with best practices as suggested by the system.
8. The system should be modifiable so that it can incorporate new and innovative modalities for clinical prediction and decision support.
9. The system should be modifiable so that important new information can be brought to the 'head of the line' under certain urgent circumstances such as drug recalls, epidemics, disasters, and acts of terrorism.
10. The system should be fully tested in prototype by expert users in parallel with the current care system before allowing it to be used in daily practice by regular clinicians. This testing will probe usability as well as detect the kinds of system errors that can only be exposed with use in a real clinical context.

Experienced clinicians make decisions with minimal or 'just enough' data- they realize that there are costs to obtaining unnecessary data. These costs include the obvious human, financial and clinical risks of further testing, but also the inevitable distractions of information overload. The ODS also introduces the opportunity for either systematic review or random auditing of clinical decisions. These audits would review system as well as individual human performance. Such analysis is already starting as organizations incorporate tools that identify clinicians who obtain insufficient, excessive or wrong data, and who make decisions identified as suboptimal under the care circumstances. Systems approaches to teaching medicine are clearly needed to prepare clinicians for optimal use of data systems.

V. Will optimal data systems improve outcomes?

First, no changes in care based on current data or processes can transcend the therapeutic limitations of current practice: Creating an optimally data driven care system is a necessary starting point in re-engineering medicine for the digital age, but it does not represent a clinical panacea. The limitations in our actual capabilities put a firm glass ceiling on the outcome improvements that can be achieved and measured without the implementation of truly innovative treatment advances. However, better use of data does provide the promise of contributing to future advances in this regard [5], and more importantly, cost-effective use of tests and treatments in the near-term. As intensivists, we recognize that critical care medicine is a particularly data rich area of medicine, but has not heretofore captured or utilized these data to a significant extent. [6]

The intensive use of data should allow us to recognize patterns in the administration of care that may contribute to otherwise undetectable positive or negative impacts on outcomes. For example, if clinicians had real-time access to prior outcomes in comparable patients, they could adjust their care plans on the basis of previously successful approaches in large populations. [3] Clinicians could also adjust their practice on the basis of observations of negative effects that can only be detected by the study of large populations. [7] These effects may be subtle or only occur under circumstances of specific combinations of clinical context and interventions, and therefore will not generally be noted in the course of normal practice or even chart reviews.

More carefully designed data presentation might speed up the process in which clinicians review data. This might simply mean better and smarter graphical displays. [8] For example, a better engineered presentation of those data elements that the clinician needs to know and can actually act upon could safely eliminate the need to review all the data entries, all the time (Figure 4).

Data can provide the basis for more robust and standardized care decisions, especially in frequently encountered situations such as acute hypotension in the ICU. [9] In addition, we may be able to use diagnostic testing in a more selective and cost-effective manner. [10] Workflow should be better supported- for example, where checklists are employed, available data could populate the checklist to some extent. Clinicians could be notified of the presence of uncompleted check list items in a manner that should improve the accuracy and diminish the tedium of the task. [11,12] Any additional provision of time provided by a carefully designed data-driven system should itself provide an advantage as clinicians recognize time as a critical limiting factor in point of care practice. [13,14]

Careful and directed use of data may allow us to discharge patients from the ICU more safely and efficiently. [15] Data can similarly be employed to identify patients with extremely poor prognoses who are receiving inevitably futile care. [16] The biggest impact of the data re-engineering is a more standardized decision-making based on predicted outcomes and retrospective comparative effectiveness analysis, avoidance of unnecessary testing, and unloading of provider cognitive workload to free up time that can be better spent on tasks that add value.

VI. Conclusion

There is always a tension between practicing optimally on the basis of current knowledge versus advancing the state of the art of patient care which requires insights and interventions not yet in the canon. This tension is the result of an unnecessary gap between research and practice: clinicians currently execute this translational process without adequate data support. Clinicians also occasionally face decisions that must be made on an individual, experiential basis, as opposed to a more standardized approach, especially when patterns have no apparent precedent in that clinician's knowledge and experience. To complicate matters, new varieties and forms of data are incrementally added to clinical databases as trials of new tests and therapies are known. The challenge to software designers and clinicians is incorporating the beneficial elements of these advances into an established information system firmly based on the integration of previously available individual and population data. Such advances will require algorithmic adjustment of the information presented to the user so that the impact of important discoveries is accelerated into a revisable and dynamically data-driven system of clinical practice.

Acknowledgments

We would like to acknowledge Ms. Kai-ou Tang who provided the figures presented in this paper.

References

1. http://www-03.ibm.com/innovation/us/watson/watson_in_healthcare.shtml (accessed 4/24/2014)
2. Steinbrook R. Improving Clinical Practice Guidelines. *JAMA Intern Med.* 2014; 174:181. [PubMed: 24322473]
- 3**. Celi LA, Zimolzak A, Stone DJ. Dynamic Clinical Data Mining: Search Engine-Based Decision Support. *J Med Internet Res.* 2014 (In press). This article describes the potential combination of EMRs with big data and search engines to provide real time decision support for clinical issues not well addressed by available clinical trial results.
- 4*. Csete M, Doyle J. Bow ties, metabolism and disease. *Trends Biotech.* 2004; 22:446–50. This work provides further engineering background for those who wish to pursue this issue in more detail.
5. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnology.* 2013; 31:1102–1110.
- 6*. Celi LA, Mark RJ, Stone DJ, et al. “Big data” in the intensive care unit: Closing the clinical data loop. *Amer J Resp Crit Care Med.* 2013; 187:1157–60. This publication provides a background summary of the data issues relevant to critical care medicine. [PubMed: 23725609]
- 7*. Ghassemi M, Marshall J, Singh N, Stone D, Celi LA. Leveraging an ICU database: Increased ICU mortality noted with SSRI use. *Chest.* 2013; 145:745–752. This provides an example of the kind of observational study to supplement and complement randomized controlled trials that can be performed using clinical databases. [PubMed: 24371841]
8. Tufte, ER. *The visual display of quantitative information.* Second. Cheshire, CT: Graphics Press; 2001.
- 9*. Mayaud L, Lai PS, Clifford GD, et al. Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension. *Crit Care Med.* 2013; 41:954–962. This is an example of the kinds of decision support tools that can be developed with more complete use of clinical data in the ICU. [PubMed: 23385106]
10. Cismondi F, Celi LA, Fialho A, et al. Reducing ICU blood draws with artificial intelligence. *Int J Med Inform.* 2013; 82:345–358. [PubMed: 23273628]

11. Vincent JL. Give your patient a fast hug (at least) once a day. *Crit Care Med*. 2005; 33:1225–1230. [PubMed: 15942334]
12. Pronovost P, Barenholtz S, Dorman T, et al. Improving communication in the ICU using daily goals. *J Crit Care*. 2003; 18:71–75. [PubMed: 12800116]
13. Cook DA, Sorensen KJ, Wilkinson JM, et al. Barriers and decisions when answering clinical questions at the point of care: a grounded theory study. *JAMA*. 2013; 173:1962–1969.
14. Westphal M. Get to the point in intensive care medicine – the sooner the better? *Crit Care*. 2013; 17(Suppl 1):S8. [PubMed: 23514562]
15. Badawi O, Breslow MJ. Readmissions and death after ICU discharge: development and validation of two predictive models. *PLoS ONE*. 2012; 7:e48758. [PubMed: 23144958]
16. Huynh TN, Klerup EC, Wiley JF, et al. The frequency and cost of treatment perceived to be futile in critical care. *JAMA*. 2013; 173:1887–1894.

Key Points

- The use of data in clinical decision making can be thought of as a clinical data system in which the responsible clinician functions as the controller.
- The current era in which EMRs are nearly universally implemented provides an opportunity for optimizing data system design to capture and leverage data in ways not available to individual practitioners in a traditional paper-based environment.
- An example of such design is real time incorporation of vast data sources into the course of clinical workflow and decision making.
- The data optimized system has the potential to improve outcomes by a variety of means such as providing useful and reliable predictions, supporting standardized approaches to clinical problems, and leveraging the data available in both population clinical databases and information resources.
- Such meticulously-designed systems will usher in a new era in clinical predictions: the interest will expand from predicting outcomes at the patient level either for prognostication or to inform decisions, to predicting information gain from diagnostic tests and response to various treatment options for individual patients.

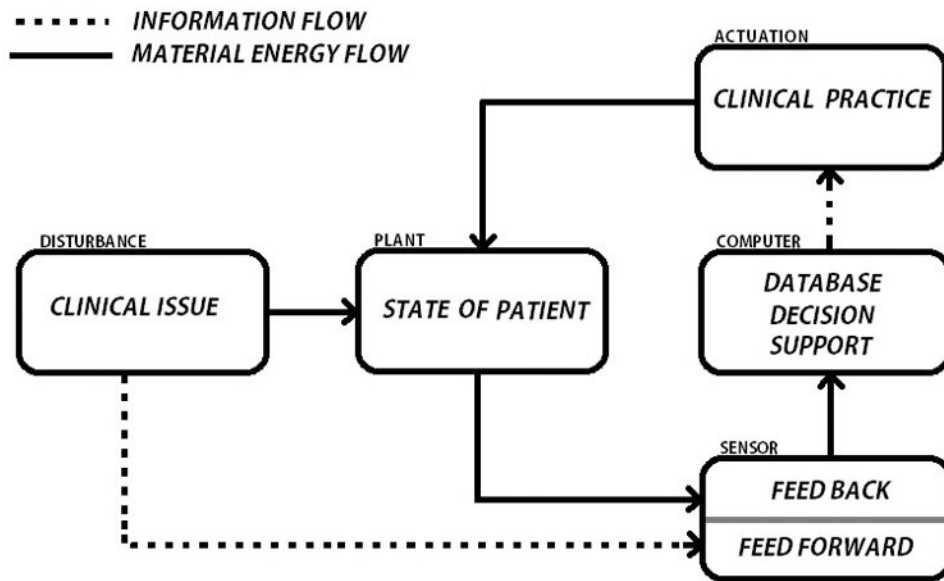


Figure 1.

Control Loop depicting a Data-driven Care system. A clinical issue such as an infection or vascular occlusion affects the state of the patient. Subsequently, the system sensor detects this change and submits the relevant data to the computer for storage and analysis. This results in actuation (or not) of a clinical practice intervention that further affects the state of the patient which feedbacks into the system for further analysis. Feedforward control involves the transmission of disturbances directly to the sensor without first affecting the state of the patient. The detection of a risk factor for venous thromboembolism that triggers prophylaxis in a protocol based manner represents a clinical example of feedforward control.

The Data Universe

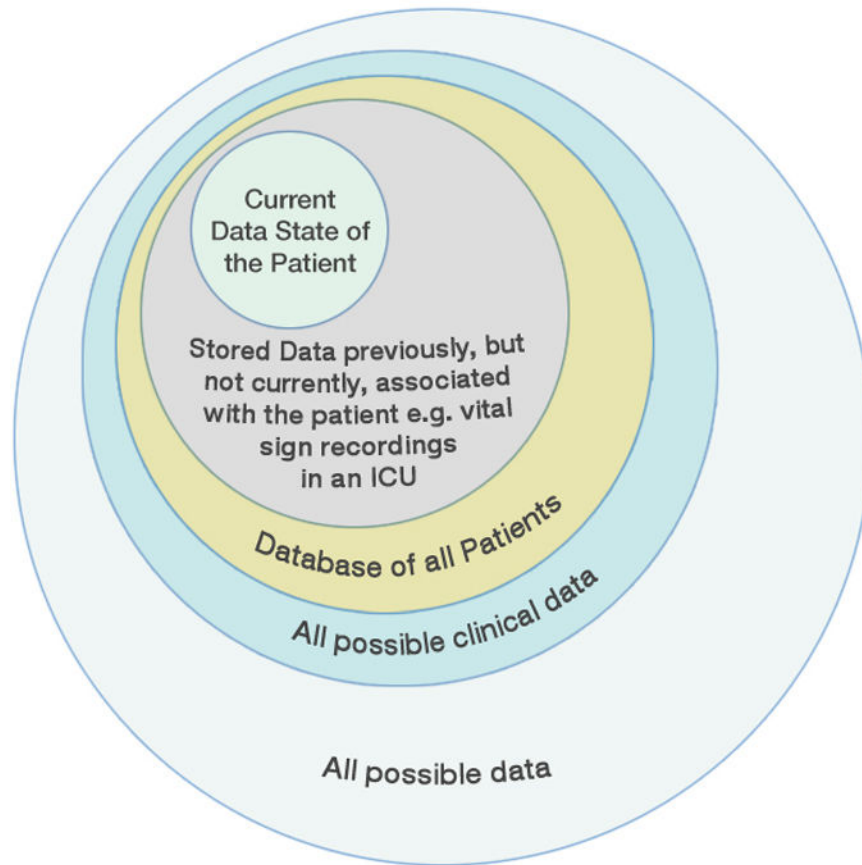


Figure 2. The Data Universe (not drawn to scale). Data move from the realm of 'all possible data' to that of 'all possible clinical data' as they are identified as having clinical value. Figure courtesy of Kai-ou Tang.

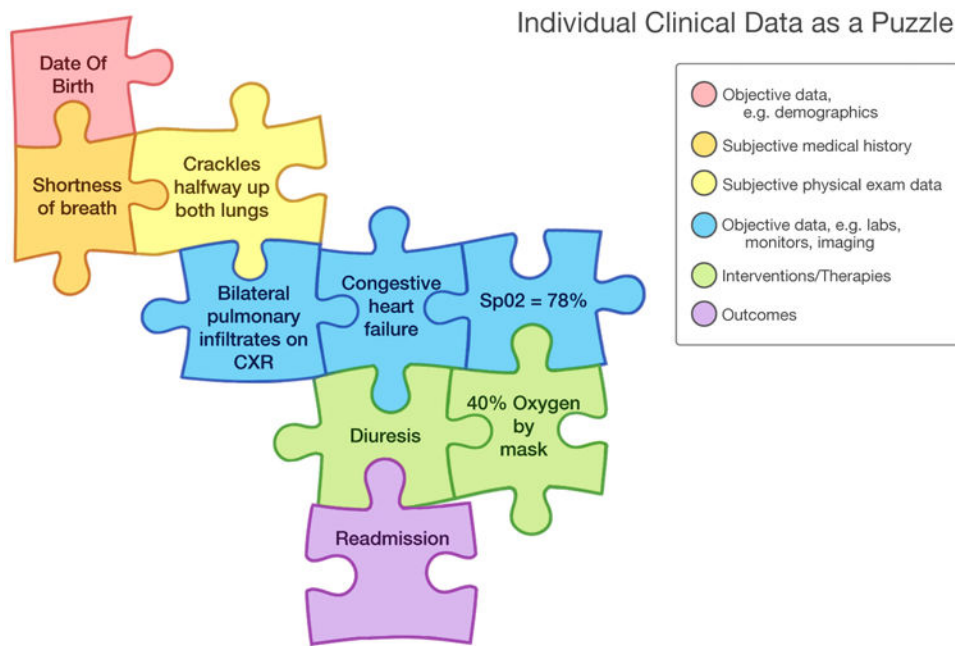


Figure 3. Individual Clinical Data as Puzzle. The puzzle changes as data are added/changed/removed but the sequence of changes can be recapitulated by virtue of date/time stamping. Decision support by population database or practice guidelines could present options for new pieces, assembly suggestions, or deletion of pieces. Figure courtesy of Kai-ou Tang.

Clinical Data Utilization

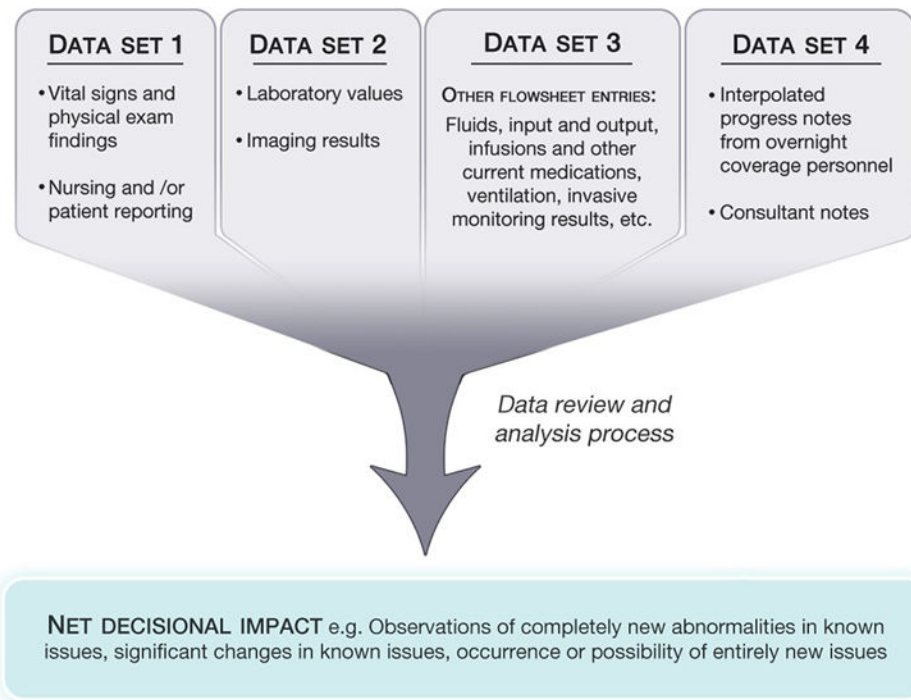


Figure 4.

Clinical Data Utilization. The clinician may analyze dozens or hundreds of individual data items in the course of workflow but only net a few significant data items that influence the course of decision making. The detection of zero change also influences the analysis. The issue raised here is how this iterative, detail-oriented process can be accelerated and supported by technology.