Bhupinder S. Juneja

Title: A Big Data Driven Systems Thinking Framework for Evidence-Informed Health Care Supply Chain Design.

Keyword(s): Health Care Supply Chain Design, Health Care Policy, Big Data Analytics, Hadoop Framework, Systems Thinking.

Introduction:

There is a strong argument to achieve a tighter integration between Health Care Policy and Health care Supply Chain Design in order to increase effectiveness and successes of health care policies and campaigns like million hearts initiative, Billion Hearts Beating, World Health Organization's Malaria eradication program, Free Polio Programs etc. Millions of dollars are spent by governments and International Institutions like World Health Organization to launch and execute these health care improvement campaigns yet it takes decades before positive interventions (if any) of these campaigns start to show up at a population level [2,11,12]. In such scenario Big Data combined with Systems Thinking framework has a potential to offer a unique opportunity to build tighter and more agile integration between public policies and supply chain by enabling us for first time in history to take an overall account of populations' needs, preferences and experience of public service [1]. Big data analytics enables us to combine heterogeneous data emanating out of both primary and secondary data sources including but not limited to formal surveys [e.g. NHANES Surveys, US Census Surveys etc] and combine it with secondary information being produced at almost real time in social media sites like facebook, Twitter etc in context of ongoing health care policy debates [e.g. some of the hottest debates on Obama Care were held on Twitter]. Together this helps us to surpass limitations posed by samples of populations and instead work directly with entire population itself. System thinking on other hand has been studied in past in context of health care Systems [9, 10] and been shown to be effective in designing more informed and inclusive health care systems. A more inclusive

and evidence informed health care supply chain design that enables us to shift focus from patient or individual driven health care to population driven health care is the key to ensure success of health care policies and campaigns.

Role of Health Care System in Society and A Case for Population Centric HealthCare.

A peek into Health care System prevalent in a society can serve as a proxy for the living conditions of the working class in a society under prevailing economic system. Study of health care can help us guide a series of economic, social and political measures that a Country must take in order to maintain stability and productivity of its workforce. Ill health can be largely attributed to social problems claimed Virchow [14]. The intent is to identify the conditions and change them so it doesn't occur. A strong argument in favor of including additional dimensions including economic, social and political in addition to patho-physiological dimension as most of studies attempt to narrowly focus on [14] in designing health care systems is that disease is multi-factorial and its treatment can't just be based on fixing a single variable that may have gone out of bounds. This in-turn helps us to scale from patient centric health care to population centric health care paradigm. A strong rationale for focusing on people centered health care emanates out of recent World Health Organization Report [6] which effectively suggests Nations to maintain a decent balance between two flavors of health care because when majority of population is suffering from epidemic disease then keeping a bunch of individuals fit won't go a long way towards maintaining the productivity and functioning of society.

Incorporating Social Priorities in Health Care Supply Chain Design.

Competition is important is a supply chain but a supply chain that fails to incorporate social priorities will eventually fail to deliver a quality health care that is accessible and

affordable to all. On the other hand a policy that focuses on affordable health care for masses [3] without providing needed incentives to health care supply chain to move in that direction will fail to achieve its intended goals. There is always a tight tussle between disparate goals of maximizing profit vs. making a service affordable to all. Such conflicting priorities are aptly documented in [3]. In this tussle, it becomes important for policy makers to include necessary incentives for supply chains to shift balance from profitability to affordability. Big data can enable to set identify the right balance between two competing choices by enabling us to statistically model demand and supply chain preferences using population level data sets instead of relying on sampling approaches. Big data can help bridge the gap between contending priorities of supply chain vs. public policy. Further it enables policy makers to seek a near real time to provide a feedback in terms of efficacy of the policies as well. In this manner it enables to generate a more refined mapping between goals of health care public policy and design of health care supply chain. Today's world is flooded with abundant amount of machine readable information available via World Wide Web (WWW) that can be accessed virtually at no cost. The www sources of such machine readable information include (but not limited to) social media sites like facebook, Twitter, Google+ to name a few, newsfeed sites, semantic web, cell phones call data records etc. Most of these mechanisms offer push and pull interfaces which became possible with advent of Web 2.0. This advent of Big data equips us for first time in History to study all dimensions of health care including social, clinical, physiological, political, economic to name a few in one setting.

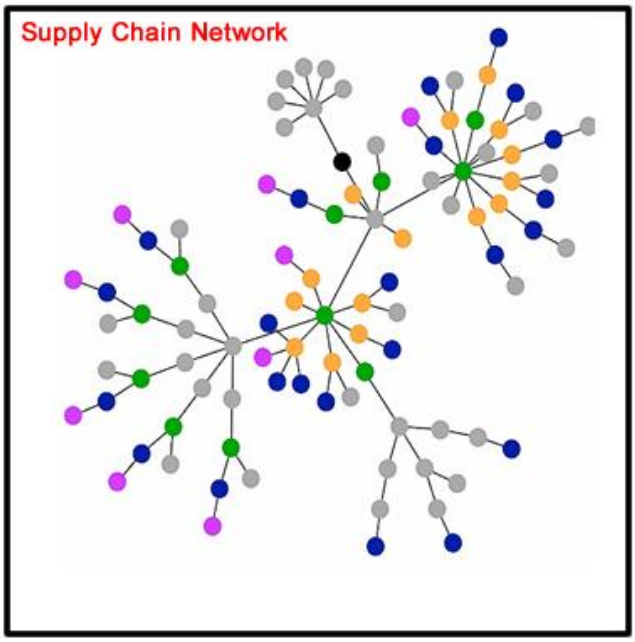
Big Data Analytics Driven Systems Thinking in Design of Health Care Supply Chain.

Systems thinking provide a useful mental construct to deal with inherent complexities of world by providing a high level abstraction of same [9, 10,15]. Systems are essentially a network

of interactions and synergies [15]. For e.g. Health care supply chain and its ecosystem of consumers, policy makers constitute one such complex adaptive system. Given its non-linear architecture and highly unpredictable behavior, It's often impossible to identify and deduce complete functioning of such system in terms of deductive mathematical formulation hence one has to often rely on inductive reasoning approaches or even atheoretical reasoning for no single theory can sufficiently explain the entire functioning of system completely or convincingly shed light on the complex system behaviors. The problem in understanding such complex system is further compounded by fact that each component (nodes and stakeholders) of system generates its own data in its own context that is very often not exactly ready for immediate consumption by other nodes on network. The section on "Nature of Data in Health Care Supply Chain Ecosystem" below sheds some light on this aspect of complex data generated by such system. System thinking provides us with a different set of tools primarily intervention (or perturbation) based set of tools to study system in which an intervention is made into the system and its diffusion effects are studied. The evaluation of such intervention in a complex adaptive system poses immense challenges owing to non-linearity of system itself and full of unpredictability. However this complexity and unpredictability lends itself very amenable to be modeled and estimated using Big Data Analytics approaches.

A Mathematical Abstraction of Supply Chain.

Supply chain is essentially a complex, multi layered network of Businesses spread all over globe characterized by ever growing competition (both between and within supply chain constituents), market uncertainty, high demand, pressure to cut costs. An overly simplified mathematical Abstraction of a globally supply chain can be attained with help of a network graph as shown below.

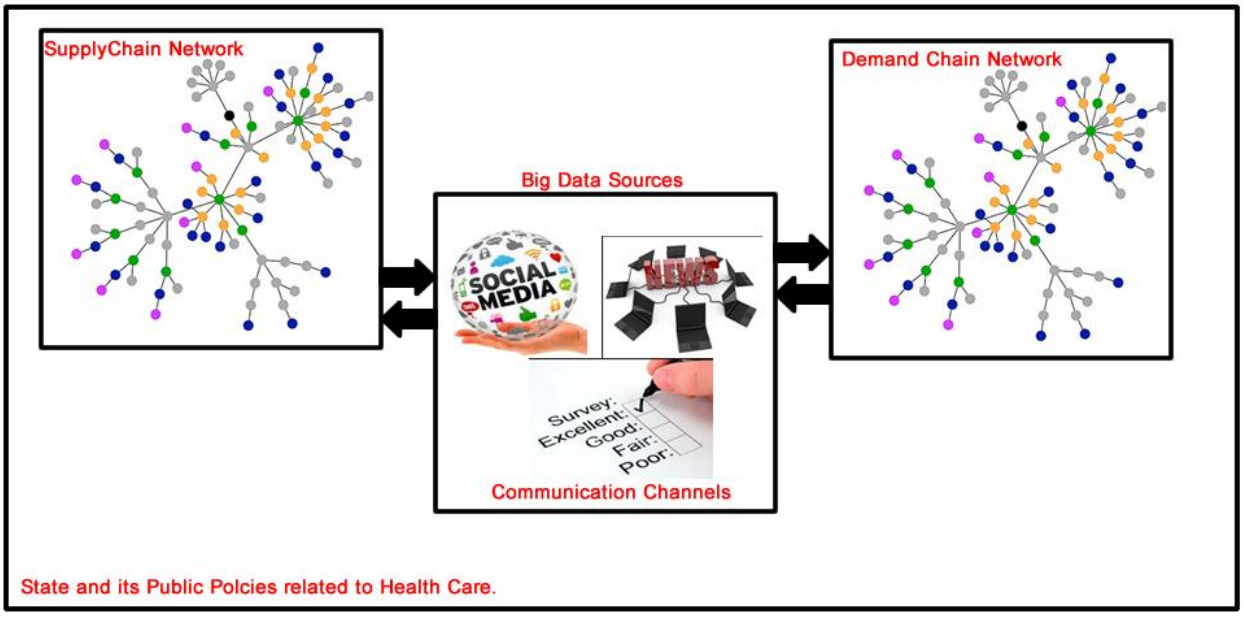


In same manner an abstraction of demand chain could be attained with similar network graph representation. The demand chain encapsulates to demand side components including patients, patient participation, accessibility and affordability of health care to patients, patients' knowledge and behavior etc [12, 13]. In summary this network based representation is a very simple model of Supply and demand chain however in reality the representation of interactions can get way more layered and complex.

On Nature of Data in Health Care Supply Chain Ecosystem.

State and its public policies are meant to balance competing interests of both supply and demand chain and enact policies that are able to represent the needs and grievances of its population yet at same time it needs to provide a conducive and thriving environment for supplies chain nodes to deliver on their performance. In this ecosystem, there are multitude of communication channels including State driven feedback collection approaches like Complex Longitudinal Surveys. These might include but not limited to US Census Surveys like Yearly

household surveys, National Health Interview Surveys (NHIS) etc. US Census conducts around 130 yearly surveys in different categories, a large subset of which could be leveraged for our purpose. Then there are CDC Surveys like National Health and Nutrition Examination Survey (NHANES) etc. Collectively these longitudinal surveys serve as primary sources of Information also known as structured data sets. On other hand secondary sources of information like Social media sites, news wires where people can provide more informal feedback on State of affairs form unstructured data sets. Apart from that there are indirect sources of information in both structured and unstructured categories that can to a reasonable manner act as proxies for providing insight into health of health care supply and demand chains and impact of policies on both. These proxy data sets could include statistics gathered by US Bureau of Labor Statistics including Inflation, Employment & Unemployment surveys, Consumer Expenditures. Collectively these could be used as proxies for well being of population and in turn help us measure the performance of health care supply chain and impact of health care public policies in improving well being of the population as a whole. The complete picture is explained with help of schematic below. Big data sources essentially attempt to combine varieties of individual level data gathered in forms of surveys and secondary data with population level data. This is where a huge potential of big data analytics lies.



Architectural Framework for Big Data Analytics :

Following sketch Borrowed as is from [8] provides a comprehensive architectural framework for Big Data Analytics.

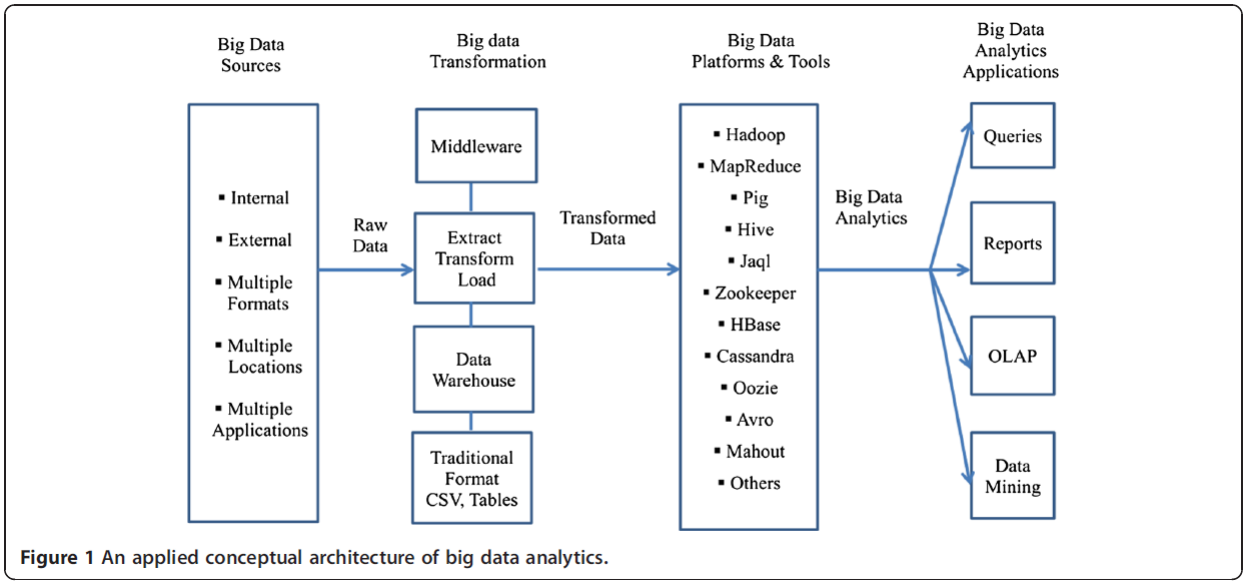


Figure 1 An applied conceptual architecture of big data analytics.

Hadoop Based Framework for Big Data Analytics Description and Key Goals.

The key tasks would revolve around combining and synthesizing of very large, complex, and heterogeneous datasets (primarily a combination of multiple survey data set and other unstructured data sets) as briefly outlined in data section above. This will then followed by the storage and querying of our combined data sets, the analysis and visualization of our data with computationally intensive models and lastly the creation of usable software for each of the tasks mentioned above. The combining of data sets and their statistical analysis will require distributed computing framework, with some tasks likely necessitating the use of MapReduce. For this purpose we will need to build an IT infrastructure (if not already available at U) that would require setting up Hadoop Infrastructure which can be done in-house or on some cloud services framework like CISCO OpenStack or Amazon EC2. On either of these cloud services we would need to set up a small (6-8 node) Hadoop cluster that contains all of Hadoop's standard components (e.g., Hive) in addition to an interface with R (including HadoopStreaming, Rhadoop, and RHive). Lastly, it needs to be clarified that entire approach would require use of a diverse set of programming languages for building software and performing analysis including but not limited to Python, C++ , R, Julia to name a few. The key goal should be to make deliverables from project as accessible to the broader research community as possible, leading us to anticipate creating Hadoop compatible APIs that could be leveraged by anyone without extensive Hadoop exposure.

General Intuition Behind Big Data Analysis Approach, A Layman's Perspective:

The key idea is NOT to start with a set of hypothesis driven by existing theories. Instead the intent is to let machine learning algorithms act on data to generate a hypothesis space for us whom we can further filter and validate with existing theory / research out there. In this manner, the data driven discovery reverses the steps in knowledge generation from statistical inference i-e

we start with data to formulate hypothesis instead of starting with hypothesis and then finding a suitable data to validate hypothesis. This reversal really forms the crux of data mining / machine learning based atheoretical approach in contrast to hypothesis driven statistical inference approaches. An elegant application of such data driven approach applied in Supply chain analysis has been documented in [16]. The key value of such approaches lies in their ability to detect rare events that are otherwise often missed in traditional hypothesis driven statistical analysis.

Potential Research Questions and Contributions:

- Can Big Data combined with Systems thinking approach help us design more informed health care supply chain and health care policies?
- Can Big Data approach with System thinking help us bridge the gap between Health care policy and supply chain competing priorities and build synergy between policy and supply chain?
- How can systems thinking combined with Big Data Analytics contribute to better evaluations of system-level Interventions in Health Care?
- Study demand side interventions via use of Social media to enhance people's participation, knowledge and behavior.
- [Methods Research Contribution]: Incorporate complex population structures underlying sampling strategies used in Longitudinal Survey data sets into existing methods for Statistical Inference. Current inference methods ignore consideration of sampling schemes. For example, standard methods may assume that observations for different individuals are independent however that's hardly a case in reality owing to

neighborhood effects and hence it may be more realistic to allow for correlated observations within clusters [17].

Conclusion:

Big data analytics enables supply chain to become more proactive in terms of its response to demand and need as opposed to reacting to same. Further Big data could help reduce waste and inefficiency prevalent in health care supply chain by almost 300$ billion claims McKinsey .Big data analytics when combined with Systems thinking can pose a great asset to build not just evidence-informed Supply chain but also evidence-informed health policies.

References:

1. http://blogs.oii.ox.ac.uk/policy/promises-threats-big-data-for-public-policy-making/

2. Bryce J et al. Countdown to 2015: tracking intervention coverage for child survival.Lancet, 2006, 368(9541):1067-1076.

3. Chilaka, M. (n.d.). Don de Savigny and Taghreed Adam (eds), Systems Thinking for Health Systems Strengthening, Alliance for Health Policy and System Research, WHO, 2009, 107 pages, ISBN 9789241563895. *The International Journal of Health Planning and Management*, 107-108.

4. Kraiselburd, S., & Yadav, P. (n.d.). Supply Chains and Global Health: An Imperative for Bringing Operations Management Scholarship into Action. *Production and Operations Management*, N/a-N/a.

5. http://www.ache.org/policy/access.cfm

6. http://www.who.int/healthsystems/topics/delivery/en/

7. http://www.hindawi.com/journals/tswj/2014/528917/

8. Raghupathi, W., Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*.

9. Rihani, S. (2002). *Complex systems theory and development practice: Understanding non-linear realities*. London: Zed Books.

10. Sarriot, E., & Kouletio, M. (2014). Advancing the application of systems thinking in health: Sustainability evaluation as learning and sense-making in a complex urban health system in Northern Bangladesh. *Health Research Policy and Systems, 12*(45), 156-176.

11. Sen, A. (1999). *Development as freedom*. New York: Knopf.

12. Travis P et al. Overcoming health-systems constraints to achieve the Millennium Development Goals. Lancet, 2004, 364(9437):900-906.

13. Victora CG et al. Co-coverage of preventive interventions and implications for child survival strategies: evidence from national surveys. Lancet, 2005, 366(9495):1460-1466.

14. Waitzkin, H. (n.d.). A Marxist View of Medical Care. *Annals of Internal Medicine,* 264-264.

15. World Health Organization. Everybody's Business: Strengthening Health Systems to Improve Health Outcomes: WHO's Framework for Action,WHO 2007.

16. Bjarnadottir, M. (2008). *Data-driven approach to health care: Applications using claims data*.

17. Skinner,C.J.(2003). Analysis of Survey Data.