The Power of Data Science for Human Health

Data Science proves its real-world value for human health from vaccine trials, to telemedicine and public health solutions, and more.

By John Doyle, Dr.P.H., M.P.H.

2020 was a difficult year for the whole world; the COVID-19 Global Pandemic caused significant strife for millions of people. However, one good thing that came out of all that tragedy was how quickly the best and brightest minds, academics, researchers, policy-makers, and pharmaceutical companies came together to produce at *warp speed* a suite of highly-effective prophylactic vaccines and cutting-edge treatments against this scourge. Myriad stakeholders collaborated in unconventional ways, forming a real-time, fit-for-purpose public health system focused on the development and distribution of vaccines and drugs.

Data Science (DS) fueled this multi-level, coordinated system, enabling this extraordinary effort and its collective success.

The massive and expeditious vaccine trials conducted by biopharma companies such as *Pfizer*, *Moderna*, and *J&J* that proved the efficacy and safety of their new vaccines represent the green shoots applications of DS healthcare.

DS has benefited human health on a number of levels from individual to community to society. Not only are we extracting increasingly robust data more rapidly from electronic health records, insurance claims, health registries, and other healthcare institutional databases; we are augmenting this data using digital health technologies such as sensors and wearables to better understand and optimize the patient experience.

This compounding clinical, humanistic, and social data is being increasingly integrated to derive insights for future vaccine and drug development. Based on these real-world insights, the potential for developing future life-saving breakthroughs is increasing exponentially.

Let's consider a few of these use cases for DS.

DS Drives Equity

How do we improve human health?

We start with creating a better experience for the patient as they interact with all aspects of the healthcare system, and we are making remarkable progress on this front—especially in integrated healthcare delivery networks. We can't stop there. Human health is ultimately governed by *social determinants of health* (SDOH) rather than hospitals, clinics, and pharmacies. Expanding access to data sets on community-level factors such as safe housing, transportation, healthy food options, education, and economic stability, are enhancing our understanding of the distribution and impact of SDOH factors.

Like any socially-driven effort, there can be many challenges along the way to building a better process-flow between patients, healthcare providers/institutions, and other community stakeholders. DS has a vital role to play in facilitating this process optimization through more efficient and effective information exchange, with the end goal of improving the patient journey on number of different levels from the individual to the population at large.

Observe the following grid:

	Individual level	Population level
Clinical Research	1	2
Real-world Research	3	4

The Level-Research Grid of DS

Number 1 involves applying DS to optimize clinical trials. Examples include, identifying appropriate patients, screening patients for eligibility, matching patients to synthetic controls arms, informing remote-based monitoring and enhancing the patient experience through digital health and tele-health medicine. In addition to speeding up trials and increasing precision of inclusion/exclusion criteria, DS can help solve for clinical research inequities by raising awareness of patient needs—especially in rural and economically-disadvantaged areas with limited access to broadband data information systems.

Number 2's DS potentiates clinical research value for biopharma, academic organizations/institutions, the federal government, and society. The premier example of this shared value creation is the aforementioned *Pfizer-BioNTech* COVID-19 vaccine trials. Without up-to-date views of COVID-19 epidemiology, along with advanced analytics predicting where the future hot spots would be, researchers would have undoubtedly taken months if not years longer to find and recruit appropriate patients for the trials. Outside of the pandemic circumstances, DS also helps triage patients for clinical research as a care option, thereby improving equitable access to the latest innovations across diverse populations. It's a win-win for all parties involved when DS is collected, processed, and applied.

Which leads to **Numbers 3 and 4**. In the real world, favorable laboratory conditions are rarely present. Consequently, the role of chance, bias, and confounding needs to be considered carefully and addressed when conducting research—especially when inferring causal relationships. This is where DS becomes indispensable and can actually guide regulatory decisions as we had seen with the temporary suspension of the *Johnson & Johnson* COVID-19 vaccine in the U.S. back in April.

Furthermore, Number 3 is about doing a better job optimizing the patient journey from prevention to long-term follow-up. While we are refining our techniques of appraising clinical risk based on genotype and omics data, we are just starting our exploration of social factor analysis considering attributes such as *health literacy*.

Much like *Amazon* product reviews and ratings, the sharing of appropriately blinded, privacy protected, and curated medical information that is matched to best represent personal interests, characteristics, and circumstances can be paramount when it comes to one's healthcare journey—deciding whether or not to get the COVID vaccine for example. And of course, this has an even greater impact when the population is trying to overcome vaccine hesitancy and reach the goal of herd immunity against COVID.

DS improve patients' experience with healthcare systems because the vaccines, drugs, and other healthcare interventions are better customized to groups by genotype, phenotype, omics, etc.

And for Number 4, researchers want to be able to study much larger, more diverse populations than the sample size in any given trial. The ability to apply data science in post-marketing research of real-world populations at large enables follow-up on safety and effectiveness across heterogeneous sub-populations and health care settings, which ultimately leads to advancing more breakthrough medicines and vaccines. As we saw with Emergency Use Authorization for the COVID-19 vaccines, accelerated approval of breakthrough medicine results in vaccines and drugs reaching the market with less data than conventional approval pathways. Often regulators require post-marketing surveillance research to follow-up on safety signals, and increasingly the real-world effectiveness is also being tested. DS plays a critical role in fueling these investigations of the benefits and risks of medicine.

DS plays a critical role in fueling these investigations of the benefits and risks of medicine.

Why is this a valuable demographic data set? Because patients in standard controlled trials are typically younger and healthier than ones in real-world scenarios. This really comes into play when researchers are trying to isolate a drug/vaccine and there's missing info when it comes to the market. Big DS, used to study massive heterogenous populations quickly and efficiently, can provide those missing pieces for say a possible COVID booster coming on the horizon.

Not just medications but innovative interventions such as recently approved surgical procedures and medical equipment are optimized by DS. Questions like 'how do the interventions work in the real world?', 'how does benefits and risk vary across diverse populations and heterogeneous healthcare settings?', and 'how are they providing value?' get answered more quickly when leveraging the power of DS.

On a government level, questions surrounding everyday drugs from prescribing, using, the most current risk:benefit and cost:benefit profile, to reimbursing help influence policies and programs such as *Medicaid* and *Medicare*.

Number 4 DS also helps us to understand both the comparative effectiveness and the cost effectiveness of new medications to market as they perform in the *real-world*. Recently the accelerated approval of *Aduhelm* (aducanumab, an amyloid beta-directed antibody indicated for the treatment of Alzheimer's disease) has precipitated a vociferous debate regarding purported benefit of the drug. Ultimately, the real-world benefit-risk profile needs to be better described, and the comparative effectiveness versus standard of care documented with the aid of big DS to settle the score. Indeed, CMS (Centers for Medicare & Medicaid Services) is considering coverage with evidence development requirements, thereby necessitating real-world studies to confirm meaningful clinical benefit.

DS can play a vital role to improve population health for making better decisions concerning drugs, diagnostic tests, surgeries, etc. Decreasing waste in the healthcare system has been a hotbed social and political issue for many years; DS can help mitigate those debates armed with objective data and facts.

The following infographic sums up how DS impacts the entire patient journey:

Prevention – Knowing population segments that are at a higher risk for various diseases either due to socio-economic conditions or genetic predispositions (like *Sickle Cell Anemia* affecting a higher proportion of minority community residents) allow for targeted community and policy healthcare interventions.

Diagnosis – High-data throughput is a result of advanced data-collection methods capable of screening entire groups of patients from infants (determining childhood cancer propensities) to COVID pool testing (for back-to-school-age students returning to the classroom). DS, in conjunction with AI-powered algorithms running on Quantum Computers can parse massive amounts of healthcare data in real time.

Treatment – Targeted treatments for specific genotypes have been recently made possible by leveraging the power of DS. Pouring through the specifics of a patient's genome to determine DNA risk factors was once the stuff of science fiction. But in today's world, this powerful treatment tool is becoming more commonplace from healthcare providers to the big biopharma.

Follow-Up – Generational and individual lifetime tracking of inherited health issues are handled by the pure application of DS. Without knowing the long-term effects of interventions, it would be impossible to know the full efficacies of drugs and other treatments.

Surveillance – And finally, without systematic surveillance of patients throughout their lifetimes, practices such as Resource Allocation would be for naught. If a group of patients (like COVID *long-haulers*) no longer need a glut of medical equipment and medicines, then that frees up those resources to be better utilized elsewhere.

Social Determinants of Health (SDOH)

Which leads to the next infographic that best describes all the challenges faced by patients every day that DS can help mitigate through better outreach (where needed), government and community policies, targeted interventions, and accessibility to life-saving drugs.

These population factors, also including *Access to Healthy Food and Transportation*, can be quantified and acted upon backed by the power of DS.

Not only do researchers measure the same things on the individual level but they're following the patient journey on a population level (i.e., community, national, and global).

This leads us back to both *Numbers 3 and 4 on the Level-Research Grid* from the previous section.

How does DS address the challenges of SDOH?

It is helping us to understand these factors, and the distribution of these factors across populations. For example, we see how these population groups interact with treatment—say in an area with poor public transportation to get to treatment centers. Also, we can see how local healthcare providers handle things like the storage and administration of medications (like the specific refrigeration requirements of the Pfizer and Moderna COVID vaccines). Other questions that can be addressed are 'can patients in a given area afford various treatment options?', or 'are there enough qualified healthcare professionals and facilities in a given area?'.

DS's value on SDOH is not only intrinsic but highly effective in communities that need resources the most.

Biopharma: Data Science is Everything

The *Pfizer* COVID-19 vaccine trials, including ages 18+, encompassed a staggering 46,331 participants worldwide from July 2020 through January 2021. The **FDA** granted Pfizer an *EUA* (i.e., Emergency Use Authorization) on December 11th, 2020. To date, over 158 million Americans have been fully vaccinated, with more than half receiving the Pfizer-BioNTech vaccine.

Consider the amount of clinical, epidemiologic, and manufacturing data needed to prove the efficacy, safety, and quality of this one vaccine to regulatory bodies such as the FDA. The public health crisis associated with the pandemic necessitated new and novel ways to collect and leverage real-world data to greatly accelerate trials without sacrificing quality. For instance, near real-time epidemiologic data describing incidence, prevalence, virulence, morbidity, and mortality were analyzed rapidly to inform inclusion/exclusion criteria and other trial design features, as well as target communities for trial recruitment based on need. In this way, real-world data science improved both the efficiency, effectiveness, and equity of trials.

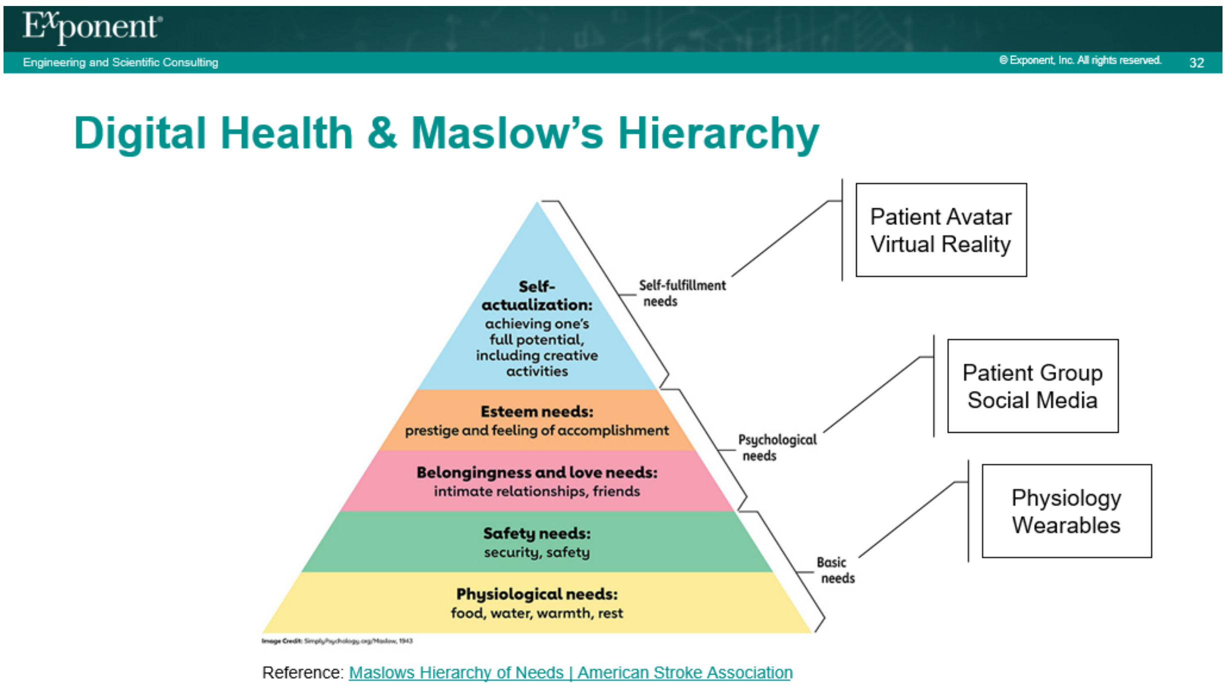
However, biopharma utilizes DS for far more than just vaccine trials (Numbers 1 and 2 on the Level-Research Grid).

DS plays a critical role in drug discovery. Advanced analytics such as artificial intelligence and machine learning are now routinely used to simulate the effects of molecules on targets in the body. This application of DS has greatly improved the efficiency of the discovery process. Deep learning methods such as convolutional neural networks are transforming the discovery process by accelerating DNA sequencing and screening biological *'omics* data efficiently for target optimization. These techniques are also helping to scan for drug effectiveness in sub-populations in prior trials in order to repurpose prior drug development for new indications, such as COVID-19 (Numbers 3 and 4 on the Level-Research Grid).

DS plays an important role in other areas, such as supply-chain resource management orchestrated with *AI* (artificial intelligence)-based demand forecasting; partnerships between biopharma organizations and academia (such as AZ and Oxford University); manufacturing operations to scale (like the staggering logistics of a billion doses of COVID-19 vaccine produced, packaged, shipped, distributed, and administered); *IIoT* (i.e., Industrial Internet of Things) leveraging the power of OT (Operational Technology) and IT (Information Technology) stacks to help produce better and safer products (e.g., consumer-grade medical devices); and training the next generation of workers in these highly-specialized fields through cutting-edge technologies such as digital thread-powered PLM, AR/VR, and spatial computing.

DS and Digital Health on the Horizon

DS and the way patient health data is collected are facing a complete Digital Transformation as new interconnected platforms become ever more available to collect patient health data for public consumption. One way to think about how digital tools impact human health is using **Maslow's Hierarchy** to frame how a patient's needs are being addressed.



Take for example wearables. A few short years ago a *Fitbit* fitness tracker was pretty much a siloed device that provided exclusive health data for the wearer such as calories burned, heart rate monitoring, and sleep tracking. However, devices like the *Apple Watch* are making headlines as the data collected from a legion of wearers (with their permission of course) are spurring not only innovation in designing better health monitoring wearables but are giving researchers new insights into population-level health habits and trends, such as the overall activity level of a wide swath of the population.

Furthermore, since the early days of the COVID-19 pandemic lockdown, there has been an exponential rise in telemedicine—reaching more patients in remote locations, while solving for equity issues as well. It's not DS on its own but the application on digital health that improves both Numbers 3 and number 4 mentioned above.

Looking at Numbers 1 – 4 on the Level-Research Grid, it becomes patently clear that Data Science is the fuel that powers digital health!

Patients First – Always

In conclusion, regardless of the use case, the application of Data Science to Human Health always begins and ends with the patient perspective in mind.

How Can Exponent Help

Exponent uniquely integrates scientific methods and engineer standards to optimize data science applications for human health. Exponent scientists and engineers bring significant experience and expertise to bear on managing risk associated with decisions made by AI systems. Whether working with structured or unstructured learning, Exponent consultants apply expert knowledge to the architecture and design of multiple layers of protection at subsystem and system levels to minimize the impact of undesirable decisions.

[Talk to an Expert]