Personalized Medicine and Whole Genome Sequencing in the Era of Big Data: Challenges and Opportunities

Gary E. Marchant, JD, PhD, Regents’ Professor of Law, Sandra Day O’Connor College of Law, Arizona State University, Tempe, AZ, USA; Gillian Bartlett-Esquifian, PhD, Associate Professor, Department of Family Medicine, McGill University, Montreal, QC, Canada; Martin Dawes, MD, FRCP, Royal Canadian Legion Professor & Head, Department of Family Practice, University of British Columbia, Vancouver, BC, Canada; and Amalia M. Issa, PhD, MPH, Professor & Chair, Department of Health Policy and Public Health & Program in Personalized Medicine & Targeted Therapeutics, University of the Sciences in Philadelphia, Philadelphia, PA, USA

**KEY POINTS**

- Widespread clinical implementation of whole genome sequencing will create massive quantities of data that will create enormous storage, analytical and communication challenges.
- The large quantities of patient data generated by whole genome sequencing, much of it of unknown or uncertain clinical significance at the time of sequencing, may expose providers to significant retrospective liability risks.
- Notwithstanding the practical, legal and ethical challenges, whole genome sequencing has the potential for major health benefits and significant long-term cost savings and represents the future of human genetics, so we need to start preparing for this implementation now.

**CHALLENGES**

Each human genome consists of slightly more than 6 billion units of just four molecules: guanine, cytosine, adenine, and thymine. The cost and time required to sequence an entire genome have dropped at an exponential pace over the past decade, and is now approaching the $1000 per genome cost target [1]. Yet, even if it is now (or soon will be) technically and economically feasible to gather whole genome sequencing (WGS) data from millions of patients, the collection, storage, and analysis of that data will present many practical, ethical, and legal complexities.

The cost estimates of WGS often focus on the actual costs of sequencing, and not many of the associated analytical and computational requirements, such as quality assessment and control, development of bioinformatics and computational tools to improve sequence analysis, management of individual sequencing projects, informatics equipment, and data analysis downstream of initial data processing (e.g., sequence assembly, sequence alignments, identifying variants, and interpretation of results) [1]. These substantial analytical, computational, and health delivery burdens give rise to the tongue-in-cheek comment that the $1000 genome will be accompanied by a $100,000 analytical requirement [2]. To add to this burden, there are already proteomics and biomarker tests now being applied to people and their tissues requiring even more analysis and interpretation.

Another practical limitation is data storage. WGS sequencing will generate terabytes of data, quickly overwhelming our current systems of storing and accessing genomic data through email and web-based databases. To make matters worse, “the cost of genome sequencing is now decreasing several times faster than the cost of storage,” suggesting that data storage is running a losing race against data generation [3]. A shift to cloud-based storage will be necessary, but that raises additional concerns about data security, privacy, and reliable access [4].

The massive amounts of data generated by WGS will also create ethical and legal risks in understanding and communicating such information. Much attention has focused already on the controversial question of what information should be disclosed to patients from WGS, but there are also important issues as to how such information will be communicated and by whom. One estimate is that explaining only the most medically significant results from the WGS of a single patient could take five hours or more [6]. If widespread WGS is undertaken, there will not be adequate trained personnel to communicate such results. For example, in the United States, there are less than 1500 clinical geneticists and less than 2500 genetic counselors in the entire nation, and these experts tended to be in short-supply even before the sequencing era commenced [7].

The deluge of genetic data from WGS is also likely to impose significant malpractice liability risks on many health care providers. There has already been a significant increase in lawsuits against physicians and other entities for mishandling genetic test results, often involving a failure to discern and communicate the clinical significance of specific genetic variants [8]. The millions of genetic variants that will be identified for each patient by WGS, most of unknown significance at the time of testing, are likely to be a landmine for physicians who will be accused, perhaps many years later and with the distorting lens of hindsight, of failing to appreciate and communicate variants disclosed by sequencing. Physicians may also face novel new liability theories with regard to their handling of WGS data, such as failure to warn patients’ relatives or failure to follow-up with new findings about specific variants.

**OPPORTUNITIES**

While the “big data” problems associated with wide-scale WGS are formidable, they are not insurmountable. The analytical and computation tools needed to process WGS data efficiently and expeditiously are being developed, perhaps more slowly than ideal, but the need and priority for such tools are increasingly recognized. Data storage technologies and management systems are also improving dramatically, spurred on by comparable big data challenges inside and outside the life sciences. New efforts are being made to educate and train health care providers to handle genetic information, as well as to provide them with the clinical support tools needed to efficiently handle sequencing information [9].

While more resources and emphasis are needed in all these areas, the key point is that WGS is coming to your clinic, ready or not, like it or not, and so it makes more sense to do everything possible to adapt to that reality now rather than to try to stall the inevitable and fall even further behind. As the cost and speed of sequencing have rapidly improved, sequencing is quickly replacing older forms of genetic testing, beginning with sequencing of specified sets of genes relevant to a particular condition (“gene panels”) and exome sequencing, but quickly jumping ahead to whole genome sequencing.

WGS sequencing is already starting to pay dividends for some patients [10]. Perhaps the primary early use of WGS is tumor mutation profiling, where the entire genome sequence of a tumor cell is compared...
to the sequence in the patient’s genotype in healthy
cells to identify mutations in the tumor cell that may
produce proteins that can be targeted by therapy.
Another early application is attempts to diagnose
patients with rare undisclosed diseases. In both of
these initial applications, WGS has achieved some
marked successes, although it does not benefit
every patient [11]. WGS is now starting to expand
into other clinical applications, such as prenatal
genetic screening, newborn screening, and healthy
adult screening to identify carrier status, disease
predispositions, and other potentially actionable
results.

A key advantage of WGS, especially in a time of
growing importance of genetic information in
health care, is that a patient can be sequenced
once and then have the data stored for life. As new
discoveries and developments occur, the patient’s
stored DNA sequence can be easily re-evaluated
on an ongoing basis for helpful new information.
With improved software (and hardware) and
our constantly increasing understanding about
genotype-phenotype linkages, re-analyzing a
patient’s genome can help identify new correlations
and risk factors. Thus, as genetics becomes more
and more pertinent to health care, an individual
born today is likely to be genetically tested as a
newborn for various conditions, genetically tested
as a young adult for serious disease predispositions,
genetically tested for carrier status prior to
reproduction, genetically tested prior to various
drug prescriptions, and perhaps also genetically
tested to compare the genotype with the sequence
of a tumor if the patient is unfortunate enough to
develop cancer. Instead of these multiple rounds of
limited genetic testing throughout the lifetime, it will
be much cheaper and more effective to sequence
the individual’s entire genome just once early
in life and then refer to that sequence whenever
useful later in life. That genomic information will
determine the number and focus of frequent
proteomic and biomic tests that occurs throughout
one’s lifetime enabling the individual to monitor the
effect of lifestyle and/or medication on the markers
for disease.

It is also important to note that health care delivery
is already complex and data-rich [12]. We are still
at a stage, however, where an individual family
doctor provides care based solely on his or her own
observations and the results of tests they selected.
This is where we will see a massive shift in medical
practice to protocol generated genetically informed
personalized care. The complexity of the data,
the complexity of the tests, the complexity of the
therapies, all require more expert systems. Today,
medical practice is like manually flying an airbus
380 without instruments so it is no surprise then
that adverse drug reactions are the fourth leading
cause of death in the United States. Perhaps the
most significant impact genetics will have is the
move from individual physician care of a patient to
individualized patient care by a physician.

It is not realistic to go backwards to a more simple
time, but rather we need to integrate new data and
technologies like WGS to improve our algorithms
to make them more accurate and effective.

CONCLUSION
It is clear that our health care system is at a pivotal
juncture in history. We have an aging population,
...the key point is that WGS is
coming to your clinic, ready or not,
like it or not, and so it makes more
sense to do everything possible to
adapt to that reality now rather than
to try to stall the inevitable and fall
even further behind.

– ADVERTISEMENT –

Oncology EMR Linked to Claims: A
Critical Link for Oncology Research

Demonstrate the value of oncology treatments using our new data
asset — linked claims data with detailed clinical data. The robust
MarketScan® Oncology EMR-Claims Data Set captures the total
patient journey — treatment decisions, outcomes, and costs.

Email us at lifesciences@truenhealth.com to learn more.
ECONOMIC EVALUATION

Modeling Alchemy: The Impact of Unorthodox Trial Design on Health Technology Appraisal Strategy

Anthony J. Hatswell, MSc, BresMed, Sheffield, UK; Marc Bardou, MD, PhD, CIC-P INSERM 1432, CHU de Dijon, Dijon, France; Meghan Gallagher, MSc, Sanofi, Cambridge, MA, USA; and Rachel Beckerman, PhD, CBPartners, New York, NY, USA

KEY POINTS

- For practical or ethical reasons, sometimes the evidence package to regulators and payers deviates from the ‘standard’ of double blind randomized controlled trials, with studies conducted without blinding, with patient crossover between arms, or without a comparator arm entirely
- The impact of these alternative trials designs on economic modeling differs, with the level of uncertainty increasing, and different techniques needed; this is particularly the case with crossover designs
- Health technology assessment (HTA) agencies are not always dismissive of non-standard trial designs, however these have caused issues with a number of products due to the nature of the evidence provided

In the past 20 years, the discipline of health economics has come a long way, with discussions settled on why modeling should be done [1] and the introduction of reference cases such that models should be comparable [2,3]. Standard approaches now exist for modeling Phase III trials, with parametric curve fitting used to extrapolate survival and other metrics, or health states used to group patients into similar ‘buckets’ to estimate rates of disease progression. There are, however, still areas with substantial uncertainty in the optimum approaches.

One such area is that of ‘unconventional’ trial design. Unconventional trial designs are usually employed in situations where there are ethical or practical concerns, leading to manufacturers having a more limited data package than that which payers have grown accustomed to reviewing: namely data from controlled, double-blind, randomized trials, with survival-based primary endpoints, and extended follow-up.

In this article, we look specifically at open-label trial designs, crossover trial designs, and uncontrolled studies as examples of ‘unconventional’ trial design. If a manufacturer is to bring a novel therapy to market with these limited evidence packages, then they may face various barriers that hinder access and pricing, including in markets where unmet medical needs remain the greatest. For each of these, we discuss how and why they arise, methods that may be employed to improve or justify their use, and the difficulties each poses for reviewers and decision makers. We also review recent medical technologies that have faced these issues, and how these have been received in health technology assessments.

OPEN-LABEL STUDIES

There are many reasons to use open-label study designs, the majority being practical issues. For instance, one treatment may be oral and another intravenous – although a double dummy may be able to be used, to do so would be onerous for patients. Another practical example can be seen with cabazitaxel (Jevtana®, sanofi-aventis groupe), which is a different color to the comparator, mitoxantrone.

When studies are not fully blinded or masked, there are a number of problems in interpretation. These are described in an excellent article by Schulz and Grimes [4], discussing issues such as the effect on compliance, concomitant treatment, and other factors. This article summarizes that from a regulatory and evidential standpoint, the more objective the endpoint, the more acceptable an open-label design – for example an open-label design using physician impression would be questioned, whilst overall survival would generally be accepted. This experience is also seen in European Medicine Agency approvals, with open-label studies generally having endpoints such as radiographic progression (done by an independent and blinded investigator) or overall survival – endpoints that are not likely to be affected by any bias from knowing treatment assignment (unlike pain score). >