

The coming era of human phenotyping

The Apple ResearchKit, the latest effort to bring deep phenotyping to the masses, heralds the arrival of large-scale human population research.

Earlier this year, tech giant Apple announced the launch of its ResearchKit. The ResearchKit is a mobile platform that taps into the iPhone's 700 million global users to find individuals interested in participating in human research studies. The first five apps included in the kit enable users to enroll in observational studies on Parkinson's (mPower app), cardiovascular health (MyHeart Counts), breast cancer (Share The Journey), asthma (Asthma Mobile Health) and diabetes (GlucoSuccess). The studies are being run in conjunction with 17 different partner research institutions and foundations, many of which are US-based. ResearchKit is the latest of several ambitious initiatives that seek to harness the convergence of mobile technology, wearable sensors for measuring phenotypic markers, and highly sensitive technologies for measuring genomic, epigenetic, proteomic and metabolic markers in blood, stool and tissue. Ultimately, the harnessing of these technologies with computational platforms to store, share, filter and analyze the data will make it possible to collect markers of health and disease data for entire human populations, opening new possibilities for biomedical research.

Human research traditionally has been based on the clinical trial. Human trials are not what one would call 'patient-friendly' activities. Gaining access depends on the competence of one's physician and his/her knowledge of the trials enrolling participants. Even if one does qualify, participating often requires a long journey to reach the hospital or facility where the trial is taking place. It can take many months before a sufficient number of people are recruited, and after trial completion, many years before analyses are finalized and the findings see the light of day in a publication: patients are often unaware of insights gained from the research in which they participated. And researchers outside the core group have little or no opportunity to access data.

This is changing with the rise of the cell phone mass market. The 1.9 billion smart phones now in use represent an unprecedented opportunity for people to connect with physicians and researchers. They potentially provide a conduit for the population to donate, follow and obtain feedback on phenotypic data they are generating—turning people from passive research *subjects* to active research *participants*.

The ResearchKit demonstrates the power of mobile technology to tap into this desire. In the first 2 days after its launch, hundreds of thousands of people downloaded, and 60,000 people enrolled in, the kit's five studies—an unprecedented recruitment rate for clinical research. Indeed, as *Nature Biotechnology* went to press, the mPower app continued to capture 500–600 people each day; 12,000 individuals were enrolled, 1,200 of which have Parkinson's. This is already one of the largest Parkinson's disease studies ever carried out. The sheer ease of connecting with large numbers of individuals through the phone will also be a powerful tool for clinicians seeking to recruit individuals into trials, particularly for rare diseases. The facility of the phone interface also means that caregivers as well as patients will be able to report data.

Another strength of platforms driven by mobile technology is that they offer the ability to monitor phenotypes in a longitudinal manner. In the case of the mPower app, trial participants track how disease affects gait, tremors, mood, cognition, fatigue, speech and sleep on a daily basis. This has the potential to reveal new patterns of markers informative about disease progression and severity. If the sample sizes are sufficiently large, there will be sufficient statistical power to find the signal in the noise. In turn, these analyses may enhance our ability to stratify diseases, which are currently defined on the basis of imprecise symptoms/phenotypes, into more finely described subgroups. Indeed, a major aim of the 100K Wellness project (currently scaling up to 1,000 individuals) the Personal Genome Project, Human Longevity and the Google Baseline Study is to compare the transition of individuals from wellness to disease. This is significant because in many cases we are unclear as to what is meant by a 'healthy person.' Finer and more exact categorization of phenotypes will mean that we will have the potential to diagnose disease faster and earlier than before, which in itself is likely to make interventions more effective.

All of which sounds great in theory. But assembling even a moderately deeply phenotyped cohort will be expensive and time-consuming.

One problem for mobile-based studies is the potential for bias. First, smart phone users tend to be from particular demographics. Second, because data are self-reported, biases arise due to potential variability in the way individuals carry out tests, the lack of rigor and fastidiousness with which they comply with study requirements and also their lack of objectivity in reporting self data. Thus, the types of observational human research currently being undertaken in the ResearchKit will need to be compared and validated with more traditional randomized, controlled trials. We will also need assurances of quality, sensitivity and accuracy for the sensors in mobile devices.

The community also needs to agree on standardized descriptions of human phenotypic variations (in a similar manner to the standardized definitions adopted for cytogenetics and sequence variation by the human genetics community). And the effort will require not only efficient ways for capturing, storing and exchanging phenotypic data (with appropriate security and privacy protections) but also new statistical models to extract features from big data. Agreement on syntactic standards will be necessary so that databases and analytical tools can understand the structure of the datasets they are given and communicate with one another, supporting robust open access on a global level. Ideally, the ability to link would go from relational databases of human phenotypic data all the way to data from animal models. Interoperability will be key so researchers can merge data sets to get the requisite statistical power.

These are major challenges, but only if they are addressed can we realize the full potential of mobile technology married to deep phenotyping and genotyping. The incentive is to bring many more thousands of participants to human research. And with greater numbers of participants will come greater power to the results.