Accelerating the Transformation of Healthcare through Open-Source Algorithms

Abstract:
Wearables have the potential to transform healthcare. They can help to inform timely, data-driven interventions. But validating each new digital measure is time-consuming, expensive, and complex. Using open-source algorithms can dramatically streamline this process. This approach creates a plug-and-play dynamic where, with reference to the V3 framework, a wearable need only be verified, if the relevant algorithm has already satisfied analytical and clinical validation. GGIR and Sleep.Py are two examples of open-source algorithms. Barriers to this approach must be overcome; adjacent opportunities pursued, so that open-source algorithms can pave a way to widespread endpoint development in a reasonable time frame and at a reasonable cost.

Today's wearable sensors can collect continuous and objective medical quality data. This data has the potential to transform healthcare from reactively treating sickness to proactively preventing it. If the technology exists, why is it not in widespread use? The short answer is that this technology must be applied very carefully and each use must be validated. For example, one study by the Mayo Clinic found that 85% of the people diagnosed with Afib by the Apple Watch did not have it.[1] Widespread use of such a tool could create much unnecessary concern among patients and significant overuse of expensive medical resources.

Adoption of open-source algorithms has the potential to dramatically streamline the validation process. The benefits of doing so accrue to both patients and healthcare companies because this approach complements – and enhances – the proprietary development of medicines and devices. To expound on this logic, we describe examples of how wearables can improve healthcare; the challenge of adopting wearables in medical applications; and the benefits of and barriers to using open-source algorithms in wearables.

Proactive Prevention Through Wearables
Wearable sensors take a variety of measurements that can potentially inform proactive intervention before the onset of more debilitating symptoms. For example, Chronic Heart Failure (CHF) is a growing problem costing $30–50 billion annually in the U.S. A wearable device can monitor a patient's activity, posture, electrocardiogram (ECG), thoracic impedance – that is, respiration and, potentially, fluid in the lungs.[2] This data can potentially identify the need for intervention before hospital readmission becomes necessary. If so, the patient's quality of life improves and their economic burden is reduced. Another even more costly example is falls: the second leading cause of global unintentional injury deaths, according to the United Nations.[3] Annually, falls cost $50 billion or more in the US alone. Yet wearables have the potential to identify people at risk for falls and appropriate exercise interventions can reduce the rate of serious falls by approximately 50%.[4]

Diabetes illustrates what can be done and the need to accelerate the process of developing these solutions. Diabetes can potentially be managed with what is effectively an artificial pancreas: a closed loop system comprising continuous glucose monitoring and an insulin pump. Steps in this direction have already improved patient outcomes and quality of life – taking us further down the path of development and real-world application than other wearable technologies.

It has taken too long however. The glucose test was developed in 1908; in 1965, the glucose test strip. In 1999, the continuous glucose monitor was developed; in 2016, the semiautomatic artificial pancreas.[5] We still await a truly automated artificial pancreas. It is clear that these technologies have the potential to improve both patient health AND healthcare costs. Why, then, does it take so long?

The Process of Adopting Wearables
Broadly speaking, the process of adopting wearables for treating disease may be broken down into 4 steps. First, you must measure something of interest (taking 2 months–2 years). Second, you must validate that this measure appropriately tracks the progression of a disease in a target population (3–4 years). Third, you must determine a clinical protocol, demonstrating what the intervention should be, when it should be triggered, what outcomes are to be expected—all optimised for patient health and healthcare economics (3–4 years). Fourth, this must be integrated into a jurisdiction's healthcare system, each one different from the next (1–2 years).

In theory, this process could take as little as 8–12 years; in practice, it is much longer. It requires a lot of work.

But could big data and artificial intelligence (AI) automate large swathes of this process, just as they have done in other areas? In reality, these techniques address only the first step. Machine learning techniques can rapidly identify unlikely connections within the data to discover measures of interest. A good data scientist can construct such models in as little as a couple of months. But collecting the data is more time-consuming and issues such as insufficient data and incomplete data also plague big data and AI techniques. Readily available data such as that from consumer wearables typically provides only processed outcomes (like step counts). Much of the information in the signal is removed and it does not have any independently verifiable data, dramatically limiting its potential benefit. Even then, AI and big data techniques do not address steps 2–4 of the process outlined opposite.

Rather than trying to build more complete, closed-loop, treatment systems like an artificial pancreas, we believe early adoption of wearables will be in clinical trials.

To be valuable for use in clinical trials, the wearable needs only assess whether the treatment being tested is working. In other words,
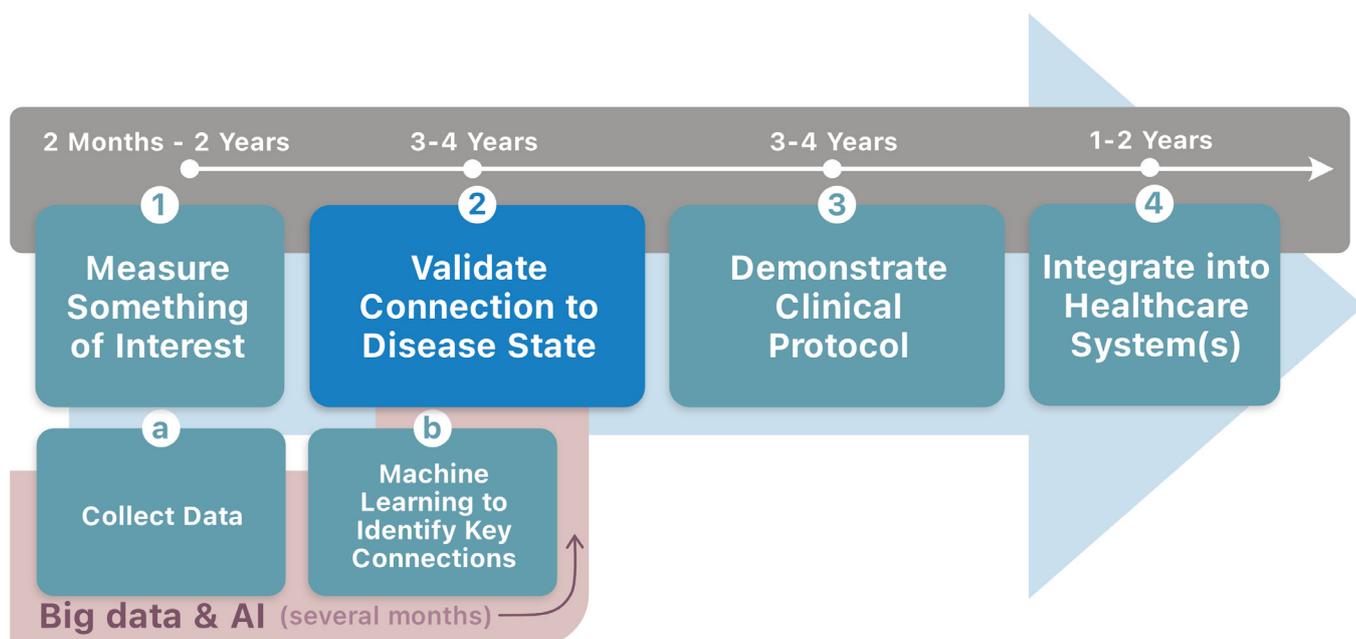
to get to the second step – validation that the algorithm is tracking the disease progression.

The use of wearables within clinical trials potentially confers many benefits, such as providing more accurate endpoints, enabling faster trials, potentially with fewer participants, and improved safety. They could also be employed in virtual or decentralised trials. Furthermore, using open-source algorithms in those wearables could dramatically streamline the progression to step 2.

The Challenges of Validation
Validating that a measure appropriately tracks the progression of a disease is, itself, a long and complex procedure. This process has been articulated using various frameworks in part or in whole. For instance, the Clinical Trials Transformation Initiative (CTTI) has helpfully set out a 13-step novel endpoint development guide.[6] Each step is accompanied by detailed guidance.

Another approach is the V3 framework championed by the Digital Medicine Society (DiMe), which is considered here. Its first component is verification: demonstrating that the wearable sensor produces accurate sample-level data. The second component is analytical validation: ensuring that the algorithms convert the sensor data into physical phenomena, such as steps. The third component is clinical validation: evaluating whether the physical phenomenon meaningfully tracks a relevant aspect of the patient's health.[7]

And herein lies the challenge: every combination of wearable, its proprietary algorithm, and resulting endpoint (or endpoints) must be qualified for each disease symptom AND each population of interest. There are hundreds of wearables, each with their own proprietary algorithms. Even a different version of an algorithm generates a different endpoint. Try conducting a simple experiment by placing two different versions of the same smartphone in your pocket. The steps measured by each device will likely be different. The possible combinations of wearables, their proprietary algorithms, and resulting endpoints, for different diseases, are in the millions. The work of validating them all is tremendous.

An Illustrative Example
Suppose the disease in question is atopic dermatitis, which involves a chronic inflammation of the skin. A wearable may produce a
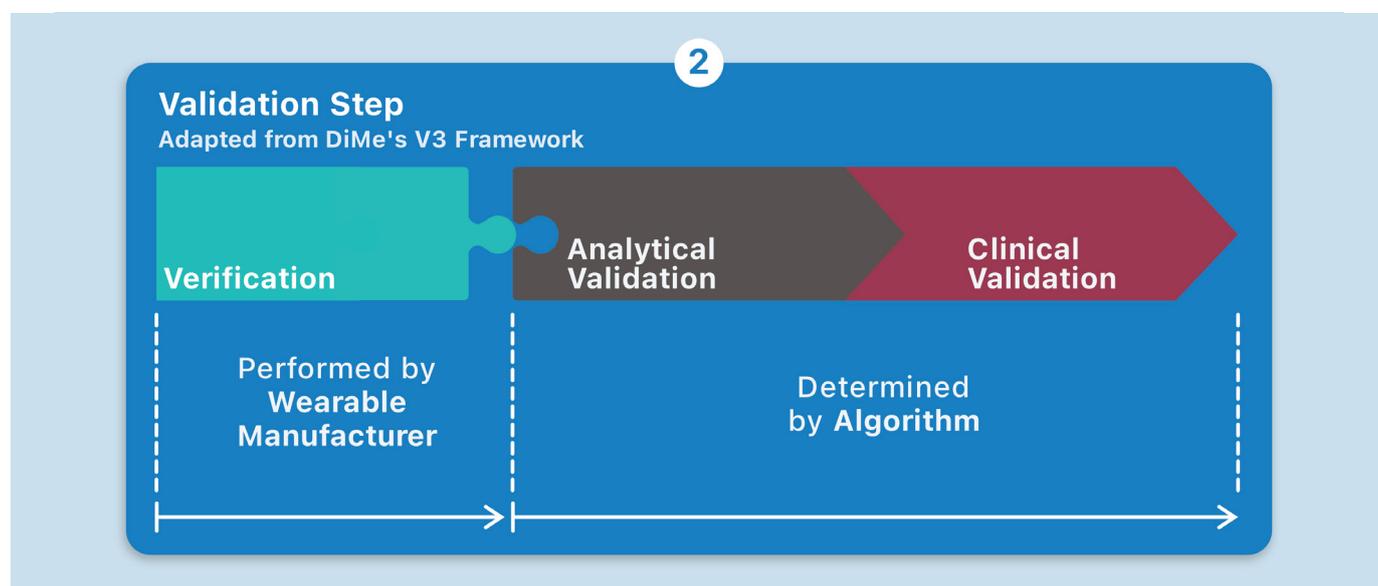
measure that tracks the progression of this disease. When evaluating the first component in the V3 framework, verification, you may demonstrate that the wearable produces accurate accelerometer data. Next, analytical validation: you then prove that its algorithm applied to the data accurately derives, say, nocturnal scratching (as opposed to a light brush, or typing on a smartphone, or some other physical phenomena). Finally, clinical validation: you evaluate the specific, algorithmically derived measure of nocturnal scratching (i.e., endpoint) for its meaningful relevance to a patient with atopic dermatitis. Only then, will you have qualified the sensor to measure this symptom of atopic dermatitis. In addition, regulatory approval must be sought. All in all, validation is an intensive undertaking. How, then, might we simplify and accelerate the adoption of wearables?

Streamlining the Validation Stage
Taking an open-source approach to algorithms for wearables can streamline the validation step without losing any of its rigor. When an algorithm is made publicly available, the inputs it requires are known. Wearable manufacturers can engineer or configure devices to produce these inputs. This establishes a plug-and-play dynamic, where the adoption of a new wearable may not necessarily entail plodding along the same analytical and clinical validation paths. If a relevant algorithm has already satisfied the analytical and clinical validation criteria, all that's left for the validation of a new digital measure is

verification, by manufacturers, to confirm that the wearable produces the right sample-level data. Transparency is key: black box algorithms diminish interoperability and hinder this dynamic.

An open-source repository of algorithms used in wearables gives healthcare companies a powerful box of tools to compare and choose from. They may also plug and play different wearables, depending on the changing needs of their study or wearable-enabled therapeutic. Existing algorithms and endpoints may be adapted to other diseases that share similar symptoms, rather than starting from a blank slate. Contributors benefit from their peers confirming the validity of their algorithms. Eventually, performant algorithms emerge. Furthermore, greater transparency and standardisation will assuage concerns of skeptical regulators down the line. The use of open-source algorithms offers a path to widespread endpoint development in a reasonable time frame and at a reasonable cost.

The Open-Source Reality
Open-source algorithms exist and are becoming more mainstream. Pfizer has been working on a nocturnal scratching algorithm, Sleep. Py, which has been made publicly available.[8] Sleep duration, sleep efficiency, activity levels and a host of other endpoints may be calculated using an algorithm within the R-package called GGIR, which was developed by Dr. Vincent van Hees. GGIR is an open-source library of algorithms for wrist-based accelerometer data. It has

been used widely in academic and other research. More than 300 peer-reviewed articles have it listed in their references. That research base is growing at more than 100 studies per year.[9] Hundreds of thousands of participants have been studied, across a wide variety of therapeutic areas and patient cohorts: cardiovascular conditions, obesity, ageing, mental health, diabetes, etc.[10] GGIR is, arguably, the most studied package of wearable algorithms available today.

Building on those advances, the Open Wearables Initiative or OWEAR was formed to facilitate the adoption of open-source algorithms used in wearables. Its website facilitates the listing of and access to algorithms and datasets.[11] Many leading pharmaceutical companies, clinical research organisations, healthcare non-profits, and research consultants are participating in OWEAR and share this vision. As with any initiative, there are barriers to implementing an open-source approach and benefits from pursuing adjacent opportunities.

Barriers to Open-Source
Open-source initiatives often provoke the concern that developers may not be sufficiently incentivised to contribute. Scientists and academics are generally rewarded for discoveries and inventions. Doing the detailed documentation and collation to make the algorithm useful to a general population requires significant time and effort and is generally not rewarded in academia. Although we have found many academics willing to share their work, they cannot afford the time and effort to make it useful.

In addition, a lack of standardisation may slow down the adoption of open-source algorithms. There is a tendency to invent new and better ways to perform the same tasks. Consequently, redundant algorithms may proliferate. But GGIR serves as an example of the tremendous benefits of having de facto standards for validation and acceptance.

Despite these barriers, a significant volume of software has been moved to open source by individuals and leading organisations like Pfizer, Mobilise-D, and Novartis.

Opportunities Beyond Open-Source
Beyond open-source algorithms, there are several opportunities that, if pursued, would make fertile ground for the adoption of wearables.

First, share data that already exists, and the task of developing and validating algorithms becomes easier. Tens of thousands of clinical trials and studies are being run. This generates vast volumes of high-quality data: detailed demographics, wearable sensor data, and outcomes data. Data sharing may, however, be met with a degree of hesitancy or hindered by competition, intellectual property, privacy and patient concerns.

Second, collect raw data because this reaps many benefits. Raw data can feed into various open-source algorithms and endpoints, is device-independent, provides a measurement of noise, enables investigation of anomalies, facilitates comparisons between studies, and leaves open the possibility of updating algorithms.

Third, the patient benefits of wearables can only be appreciated with a paradigm shift from population thresholds and averages to self-benchmarking. At Shimmer, we have conducted numerous studies using participants' galvanic skin response (GSR), or changes in sweat gland activity, to measure their emotional response to different stimuli. These studies were conducted in laboratory conditions – so the data was as clean as one can get. Yet participants' baselines may vary by a factor of 40 or more in the same study. Different people also

have different response levels. And despite the excellent conditions, there was still quite a bit of noise in the data. These observations are commonly found in other wearable sensor data. A standard approach of setting ranges and thresholds based on population norms and averages may not work. However, we can make sense of the data by benchmarking individuals against themselves and determining whether they are responding. Then we can calculate the percentage of participants responding at any given moment, perhaps, in response to stimuli or treatment.

Summary
We live in a world of wearables. In healthcare, wearables and other digital technologies have the potential to inform proactive, more timely interventions, improving patient quality of life and reducing costs. But the process of adopting wearables takes time. Validating measures produced by wearables for medical applications is particularly rigorous and complex. An open-source approach to algorithms used in wearables streamlines the validation stage. It establishes a plug-and-play dynamic—such that wearables need only be verified, if the relevant algorithm has already satisfied analytical and clinical validation, to produce a validated digital measure. Initiatives, such as OWEAR, already exist to foster such collaboration around endpoints and algorithms. The GGIR library is a foretaste of what can be done. Barriers to the adoption of an open-source approach need to be overcome, so that wearables in healthcare can become an everyday reality.

REFERENCES

1. https://www.mobihealthnews.com/news/apple-watchs-abnormal-pulse-feature-driving-many-unnecessary-healthcare-visits-mayo-clinic, visited 31 Jan 2022.
2. Shimmer internal analysis
3. https://www.who.int/news-room/fact-sheets/detail/falls, visited 31 Jan 2022
4. El-Khoury, F., et al. The effect of fall prevention exercise programmes on fall induced injuries in community dwelling older adults: systematic review and meta-analysis of randomised controlled trials. BMj, 347 (2013).
5. Hirsch IB. Introduction: History of Glucose Monitoring. Role of Continuous Glucose Monitoring in Diabetes Treatment. Arlington (VA): American Diabetes Association, 2018 Aug.
6. https://ctti-clinicaltrials.org/wp-content/uploads/2021/06/CTTI_Novel_Endpoints_Detailed_Steps.pdf, visited 31 Jan 2022
7. https://www.dimesociety.org/tours-of-duty/v3/, visited 31 Jan 2022
8. Mahadevan, N et al. Development of digital measures for nighttime scratch and sleep using wrist-worn wearable devices. NPJ digital medicine, 4(1), 1-10 (2021).
9. https://www.owear.org/ggir-airtabl, visited 31 Jan 2022
10. Migueles JH, Rowlands AV, et al. GGIR: A Research Community-Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. Journal for the Measurement of Physical Behaviour. 2(3) (2019).
11. https://www.owear.org/open-algorithms, visited 31 Jan 2022

Geoffrey Gill

Geoffrey Gill, MS, is President of Shimmer Americas, leading the U.S. operations and the commercial efforts for North and South America for Shimmer Research, a designer and manufacturer of medical-grade wearables. Geoffrey is also a Co-founder of the Open Wearables Initiative (OWEAR), an industry collaboration designed to promote the effective use of high-quality, sensor-generated measures of health in clinical research through the open sharing of algorithms and datasets.