The Digital Twin in Clinical Research

In industrial applications the concept of digital twins is successfully applied currently to gain insights into complex processes. In clinical research this concept would offer impressive advantages. Its implementation requires major challenges to be met.

Background

Clinical research is facing a paradoxical situation: Production processes in the pharmaceutical and medical device industry are driven by a high level of automation, control, and traceability. Digital twins of assets and processes are established by using internet of things (IoT) technology to acquire massive amounts of data in real time. On the other hand, clinical research still relies on collecting huge amounts of data from subjects at rather few points in time. It relies on data being manually entered into CRFs or being manually copied from medical records. Things turn out to be even more paradoxical, when the same subjects are using wearable devices to continuously track sleep, heart rates, or physical activity. They are creating their digital twin as well. Indeed, it is a paradoxical situation. But creating and using digital twins in clinical research could boost efficiency and data quality.

Digital Twins in Industrial Production

In industrial production, the concept of digital twins is gaining a lot of attention throughout the past few years; it is thought to be one of the major trends in information technology as well. In industrial production, such digital twins are the key to real-time optimisation and decisions. They are used for simulation, for education, and for virtual prototyping as well. Some of the most interesting applications can be found in the field of predictive analytics, when digital twins are being used to predict future performance and failure of their counterparts.

Being a digital, thus virtual, representation of assets, processes, people, systems, etc., a digital twin needs to be far more than a structural design. They need to include dynamic properties as well. Such properties are derived from sensor data that is being acquired in real time, which is probably one of the most important applications of so-called internet of things (IoT) technology. Following this concept, the digital twin's physical counterpart is equipped with a set of connected sensors that collect data continuously. A digital twin therefore basically consists of a huge and growing amount of data that sufficiently reflects all required structural and dynamic properties.

Of course, the massive amount of data, all related hardware and software, as well as the corresponding statistical approaches brought a revolution to industrial production, which requires the convergence of operation technology (OT) and information technology (IT). Decreased lot sizes, reduced maintenance efforts, and improved quality are promising results of current digital twin approaches.

The Visit Gap

At first glance, clinical research and production might look rather disparate. A closer look unveils important similarities:

- In both worlds a lot of effort is put into ensuring a high level of quality and traceability.
- Processes are usually standardised and documented, which allows for automation and modelling.
- Collecting, storing and analysing data is inevitable.

When clinical data is being collected through the course of a study, this aims at providing sufficient evidence for efficacy, tolerability, and safety. Such data therefore serves as the subject's digital twin. However, the way data is acquired for this digital twin does not fulfill the requirements that we would expect nowadays for a digital twin in production, because clinical data is usually collected at very sparse points in time, usually during the subject's visits (Figure 1). This visit gap in data acquisition might be appropriate for slowly changing parameters like the subject's weight. Fast-changing data (e.g. physical activity) cannot be captured adequately. Spontaneous events (e.g. cardiac arrhythmia) might simply be missed, if they do not occur during a visit. The visit gap therefore imposes major limitations to the digital twin, which is why the gap should be narrowed down.

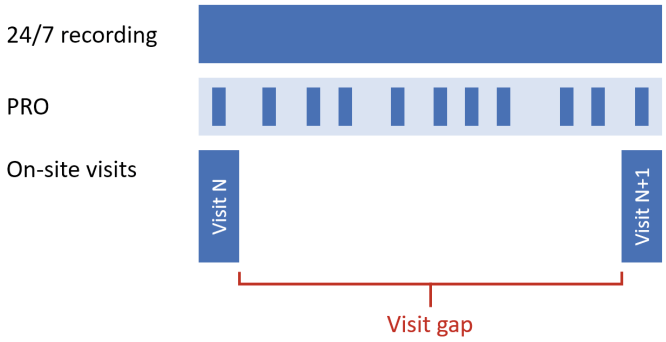

Figure 1. Data acquisition in presence of visit gap

Bridging the Visit Gap

Traditionally, the visit gap in data recording can be bridged by using patient-reported outcomes (PRO), either with paper-based diaries or with modern ePRO software. PRO approaches clearly rely on the subject's compliance in entering the correct values at the intended time or event. Therefore, a lot of effort is put into monitoring the subject's compliance. On the other hand, using PRO implies parameters that can be recorded by the subject.

The PRO market has already seen several diagnostic or therapeutic devices that have been extended by diary functionality, reminders etc., e.g. for COPD monitoring. Such devices can improve compliance in using the devices and entering data. However, they still require interaction by the user and do not usually record data in a continuous way.

Continuous data recording would ultimately bridge the visit gap, because relevant parameters are recorded 24/7. For such purposes, wearable devices (Figure 2) are spreading in the consumer market, where users try to learn more about their individual level of fitness. Most of the time those wearables are worn as smartwatches or integrated into textiles. They provide continuous recording of heart rate, ECG, blood pressure, blood glucose levels, physical activity,

temperature, sleep, physical stress, and much more. Wearable technology brings in many important advantages for within clinical research:

- Wearables can be used at home requiring only few training efforts.
- Data acquisition is performed automatically, i.e. the subject is not required to enter data or to trigger recording manually.
- Automatic recording reduces effects due to subject non-compliance and potential fraud.

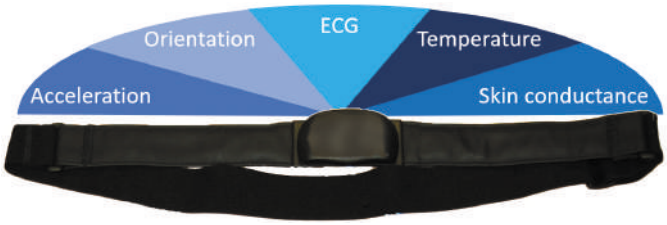

Figure 2. Example of a wearable device worn as chest strap

Rethinking Data Analysis

Diaries and CRFs are designed to collect data in a structured and study-specific way. Wearables and other network-enabled diagnostic or therapeutic devices often provide less structured, complex streams of real-time data. Such data streams offer deep insights into the subject's health status as well as treatment effects or safety events. Unveiling such relevant pieces of information from a vast amount of data requires the use of advanced data analysis techniques.

Traditionally, activity measures and quality of life measures are highly aggregated scores that are provided by the subjects themselves. With 24/7 recordings of physical activity such scores can be derived in a more objective way (Figure 3). The same applies to stress levels, sleep, etc. Safety-related events (e.g. arrythmia, immobility, fever, etc.) could also be detected in the data streams automatically, which already is applied in consumer devices.

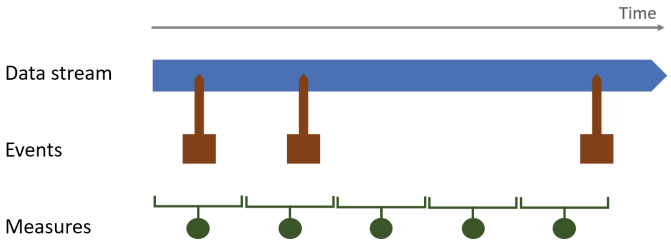

Figure 3. Processing of data streams

Technical Challenges

Applying analysis techniques to streams of data might sound straightforward; in fact it assumes a single stream of data to be available. Most of the time, multiple sensors are needed; their individual data streams need to be concentrated by gateway systems, which requires common communication protocols to be used for every sensor. Unfortunately, the market for wearables and for other diagnostic and therapeutic devices is still rather heterogenous towards communication protocols.

The massive amount of data requires a rethinking of data storage. Database technology, especially so-called timeseries databases, is required on the device level. However, there are important strategic decisions needed: Uploading all data to cloud systems could result in data privacy issues and cause high archiving efforts throughout the retention period. Therefore, data could be pre-processed locally and only relevant features (e.g. time and characteristics of a cardiac arrythmia event) be uploaded to the clinical database, while irrelevant data is omitted automatically. This approach is currently being used in medical event recorders as well as in consumer-oriented wearables.

Besides those rather strategic questions, there are some rather practical considerations:

- Low power consumption and user-friendly charging options (e.g. wireless charging) are needed for battery-powered devices.
- Broadband communication might be unavailable due to the building's structure or to regional restrictions in service.
- Firmware upgrades need to be applied without user interaction.
- Technical support needs to be available for a variety of devices.

Data Security and Privacy

Data security traditionally focusses on confidentiality, integrity and availability. From the technological point-of-view there are well-proven approaches to ensure security for the digital twin and the devices used to maintain the digital twin. Nevertheless, the scenario involves some threats that are challenging in total. All devices are connected through networks, most of these being wireless networks. Every device itself could be compromised, which would affect the whole network. Threat modelling and security analysis is inevitable, as are secure programming techniques. The variety and rather large number of devices involved requires automatic monitoring systems to be in place that detect outdated firmware revisions, manipulation attempts or limitations of service.

Privacy tends to be critical for the subject's digital twin, because it offers deep insights into the subject's health status. Therefore, common privacy standards, like the GDPR, impose additional requirements for personal health information (PHI). Those include data encryption, access control, the individual's control, etc. PHI becomes even more critical when 24/7 data streams are stored, because they would allow insights into the health status that exceeds the trial's focus, which raises ethical questions as well. In any case, the digital twin must include a sponsor view that ensures necessary data and information are hidden.

Regulatory Challenges

Decentralised 24/7 recording and storage of data raises the question of source data and audit-trail maintenance. The same applies to raw sensor data versus processed event data. Consumer-oriented device manufacturers are struggling with the national certifications, because of the different regulations. Those challenges are accompanied by discussions related to direct data capture (DDC) that we have seen during recent years. All those questions involve a high level of uncertainty, which is also reflected by contrary opinions of auditors and inspectors.

At first glance, validation activities might look much easier to handle than the rather general questions raised above. A lot of experience is available towards device validation. However, there are specific characteristics that need to be considered, e.g. security testing, distributed systems, usability, and most of all, assessment of performance and reliability for powerful data analysis techniques.

From Subjects to Trials

Besides the subject's digital twin, there are other applications where digital twins could improve processes in clinical research. Every clinical trial itself could be digitally represented by a digital twin. This twin would comprise the trial's structure, i.e. visits, inclusion/exclusion criteria etc., as well as dynamic properties. Data needed for this can be acquired from various electronic tools available,

e.g. e-CRF, e-TMF, IWRS, CTMS etc. Merging this data into a digital twin appears to be quite similar to current approaches like dashboards. Its intention, however, is different in a way that it is less focused on specific KPIs or rule-based alerts, but rather provides a rich data source for anomaly detection. Where KPIs and rules are used to monitor expected deviations, anomaly detection can unveil unexpected deviations upon statistical measures.

With the traditional set of tools and processes, a trial's digital twin could suffer from some data being collected only during on-site visits. This can best be seen when looking at the several logs and trackers that the monitors need to check. They also indicate potential for acquiring data automatically. IoT technology, e.g. identification techniques, connected devices, etc., can provide such data. Moreover, integrating all subjects' digital twins into the trial's digital twin could massively improve availability and completeness of data. This would change monitoring and oversight processes completely.

Conclusion

Digital twins in clinical research offer impressive advantages like 24/7 data recording, automatic assessment of health status, reporting of safety-related events, etc. Appropriate analysis techniques need to unveil relevant pieces of information out of heterogenous streams of real-time data, which could be acquired by using IoT technology. Wearables are a well-proven approach for 24/7 automatic recording of subject-related data. IoT technology as well as the corresponding storage, communication, and analysis processes are spreading outside clinical research. While the technology basically is available, there are major challenges related to data security, data privacy, and regulatory compliance. Those challenges and concerns might remind us of the change from paper CRFs to electronic CRFs. Besides such challenges, there is the everyday challenge of recruiting subjects. Digital twins can lower the burden for subjects to fulfill their duties because of automatic data recording. Of course, if subjects are used to track their personal health status with smartwatches, connected balances, etc., they would expect clinical research to use such technology as well.

Jörg Mielebacher

Jörg Mielebacher (PhD) has studied medical informatics and has worked in this field for more than 15 years. He is an experienced IT consultant and software specialist, with a specific focus on automated analysis of processes in clinical research. He is the founder and CEO of PCQ Pilots in Bremen (Germany), a vendor of various software solutions for clinical research, including e-CRF, e-TMF, risk management, randomization, and training. He is also an honorary professor for computer science.

Email: joerg.mielebacher@pcq-pilots.com