5. Fit-for-purpose clinical data strategies (the 5 Vs)

The evolution of clinical research practices and supporting regulations, as well as massive advances in technology have fundamentally changed what clinical data is. As we define the future beyond traditional EDC, we need to rethink our approaches and understand how the "5 Vs" of clinical data are re-shaping CDM.

5.1 Volume (Terabytes, petabytes, exabytes to yottabytes)

In 2012, Tufts[3] estimated that in average, phase III studies collected close to 1 million data points. Today we measure m-Health data points in the billions. This demands the adoption of new strategies to improve the collection, processing and archiving of data supporting this new scale. CDM must re-imagine its practices to efficiently move **from a few datapoints per CRF to more than tens of thousands of datapoints generated per patient per week**.

Figure 2 shows the expected volume of actigraphy data generated by wearables (in blue) compared to data generated from site visits (in orange). The protocol requires 260 patients to be treated for 6 months. The enrollment period is estimated to last 6 months. With wearable device set to transmit data every minute, wearables would generate a pulse reading more than 68 million times. In comparison, pulse would only be generated 3,380 times through site visits, assuming patient's visits every 2 weeks.

With the incredible increase in data volume, CDM must be diligent and secure Quality by Design (QbD) by defining what really needs to be collected to support the protocol hypothesis vs. all data that can be generated through new technologies. Not all data generated by devices may be useful for statistical or further exploratory analysis. In the case of wearables, CDM may consider retaining the 68 million pulse readings as e-Source data while only retrieving data summaries at regular intervals (e.g., every hour or day). Data collected may only include key data characteristics (e.g., min, max, average, standard deviation, number of observations generated, etc.), aggregated (e.g., by time reference such as epoch) to better support downstream activities such as safety monitoring, data review and statistical analysis.

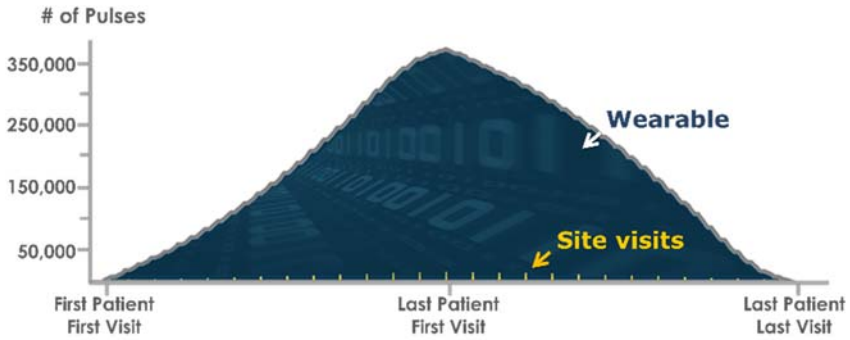

Fig 2. Daily volume of actigraphy data from wearable vs. e-CRF pulse data

Another example of data expansion is the rapid increase in focus on wellbeing, and the array of passive data sources now being made available for research. According to the IQVIA[4] Institute for Human Data Science, there are currently over 318,000 health apps and more than 340 consumer wearable devices tracking, measuring, monitoring and connecting healthcare stakeholders. Additionally, there are more than 200 new health apps added to app stores every day.

It is therefore not surprising that sponsors are increasingly using digital health technologies in clinical research and leveraging apps to collect reported outcomes and other real-world data (RWD). However, most experiments with digital health have been confined to Phase IV trials, reflecting the perceived risk of incorporating digital measures into pivotal trials until they are validated and pressure tested.

This is unfortunate as those technologies can improve the efficiency of clinical research in many ways. Solutions for identifying sites, targeting and recruiting the right patients, collecting reported outcomes, gaining digital consent, screening patients remotely and conducting decentralized trials have all proven to be effective and useful. First and foremost, they benefit patients by removing enrollment barriers and enabling breakthrough medical advances especially for rare diseases. Solutions such as telemedicine can benefit sponsors by reducing resources, optimizing site selection, speeding up enrollment, easing data collection and supporting rapid decision-making through immediate access to data. Additionally, biosensors, predictive analytics and novel patient assessment media are leading to new discoveries, reducing time to insight and optimizing patient identification. The FDA[5] has released extensive guidance on the appropriate use of such technologies, as well as developing new endpoints vs. establishing equivalence with existing endpoints. To achieve our CDS goal of managing this growing volume of data, we must develop new data science principles, data collection tools and data analytics strategies.