

Paper RW03

Early experiences using SDTM to organize clinical data collected using a physical activity tracker

Martin Gram, Novo Nordisk A/S, Copenhagen, Denmark
Gianluca Mortari, Novo Nordisk A/S, Copenhagen, Denmark

ABSTRACT

Real World Evidence provides pharmaceutical companies with opportunities and insights about their drugs in ways they haven't had before. Wearable technologies play a growing role in Real World Data generation and there is a large potential to pursue novel endpoints that has clinical relevance within many different disease areas. This type of data is unlike traditional clinical data both in terms of increased data volume and the inherent nature of the data. At the same time standards for physical activity data is relatively immature. This paper will share some of our experiences and learnings we have gathered while preparing our first generation of SDTM data based on an FDA class II wrist worn activity tracker. In the first part, we will discuss challenges with potential endpoint definitions for physical activity, data collection and data handling decisions. In the second part, we describe the choice of SDTM domains and the impact of data volume on the SDTM data structure. At the end of each paragraph is a summary question where we believe additional guidance from standard development organisations and authorities are warranted.

INTRODUCTION

The physical function of subjects and possible changes with treatment has traditionally been evaluated with questionnaires or functional tests such as the 6-minute walk test. However, these tests have a number of limitations and objective measurements over a longer period can provide a more representative result and a better insight into possible treatment effects. We identified the opportunity to use a wearable physical activity tracker to characterize the physical activity level in a sample of individuals in a study. As activity trackers, accelerometers and pedometers are becoming common in use, the threshold of adopting these techniques into a study may well be low. However, digging deeper into the various aspects of data originating from a wearable device we encountered some challenges and learnings which form the topics of the present paper.

FIRST PART

KNOW THE DATA

Unlike traditional clinical data where the source data has a meaning by itself, physical activity data carries no meaning and is commonly not used directly as an endpoint. The raw data output from an activity tracker is the accurate biomechanical representation of movement and the acceleration measured is therefore a complex mixture of the internal and external forces generated. This could be human movement, but also external noise generated from equipment like lawnmowers or handheld tools. Thus, the association between raw acceleration measured and physical activity (usually metabolic cost) is rather complex. If the raw acceleration is aggregated into a physical activity (PA) metric it is possible to demonstrate a simple linear association between the metric and the metabolic cost for locomotion activities (1, 2). For this purpose, algorithms are made which associate the raw data output with a meaningful measure of physical activity using advanced machine learning methods like neural networks (3). There are many different ways in which un-processed acceleration data (raw data) can be summarised to create relevant PA summary statistics (4), however, currently there is a lack of consensus on what methods provide accurate results (5).

SELECT THE RIGHT DEVICE

In general, wearable devices can be divided into two categories; consumer grade wearables and research grade wearables. Most consumer grade activity trackers provide a variety of endpoints associated with PA. This could for instance be pedometers, where the most basic measure is the number of steps over a period of time. Common for these devices are the fact that the algorithms applied are proprietary and the validation of them not publicly available. This makes them unsuitable for clinical trials as the data transparency is low and the validity of the data is unknown.

Research grade instruments on the other hand store un-processed PA data i.e. acceleration, which can be used as input to publicly available scientifically validated algorithms which then create the endpoints. In order to submit data to FDA, the devices used must be classified as class I, II or III to fulfil the requirements for submitting collected SDTM data. Earlier it was a challenge that wearable devices stored all the data on the unit itself. This has previously led to a substantial number of episodes with missing data as the devices were lost before data upload could proceed. We recommend pursuing newer solutions which allow daily data uploads via Bluetooth to the cloud when the subject is near a hub situated at their own home. Furthermore, the battery capacity and the robustness of the device towards dust and moist are other important aspects of the device to ensure minimal missing data and a good user experience. As these aspects are likely to be unknown territory for most pharmaceutical companies we suggest doing research early in the project to identify and partner up with an appropriate vendor with the required knowledge and logistical setup.

CHOOSE THE ANATOMICAL WEAR LOCATION

Different anatomical wear locations are possible with wearable devices (i.e. hip, ankle, lower back, wrist, waist etc). The wrist placement has over the last few years gained widespread use motivated by the increased wear comfort, which also improves wear compliance (6). High compliance rates are crucial for the data quality and low wear compliance can lead to selection bias and misclassification (7). The wrist placement is used with large-scale surveillance studies like NHANES (US) and UK Biobank (8, 9). It is easy to wear on the wrist and interactions with subjects on challenges that occur while wearing the device is easier compared to other anatomical locations. However, historically the hip has been the preferred anatomical location. This seems to be motivated by the fact that hip movement is associated with the movement of the centre of gravity and thus represents gross whole-body movements and a relatively tight relation between objectively measured PA and energy expenditure compared to the wrist placement. In the current literature, wrist placement has generally been found to be less accurate than waist placement in PA behaviour classification (10-13). The largest drawback with wrist placement is the difficulty of current algorithms to detect sedentary time as fidgeting resembles slow walking. Nonetheless, the number and quality of wrist specific algorithms will likely increase substantially in the near future as a result of machine learning techniques. This prediction is partly based on the fact that the development of devices capable of measuring PA for private costumers have almost exclusively focused on wrist worn devices.

The question is which anatomical location is considered valid in a submission?

NON-WEAR, SLEEP AND IMPUTATION

Non-wear are periods during a day where the subject is not wearing the device. Zero activity is recorded during this period although the subject likely did some activity. It follows that any non-wear during waking hours will underestimate the total daily activity level. The level of underestimation is dependent on the time of day that it happened because the physical activity level fluctuates over a day. To avoid non-wear most protocols, require the subjects to wear the activity tracker 24 hours per day. The amount of non-wear is historically in the order of 2-4 hours per day, lowest when using the wrist location, but it also depends on the visit structure of the study i.e. if feedback on wear compliance is allowed or feasible (14). In an article from 2017 (15) Schneller et al wrote: "...non-wear leads to arbitrary decisions regarding wear time validation criteria and inclusion criteria for a valid day, which can significantly affect the compliance rates (6), wear time (16), time spent in different PA intensities (17), and differential bias by exclusion of participants with certain characteristics (18). Although non-wear must be held to a minimum, in practice non-wear will always occur.

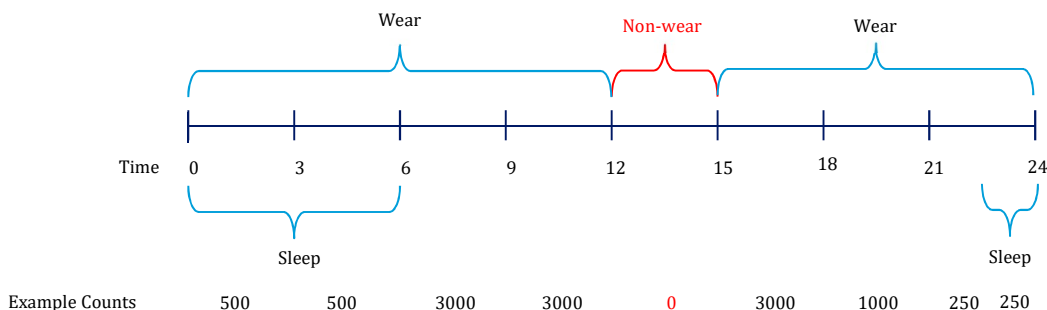


Figure 1. The figure shows an illustration of the distribution of sleep and non-wear during a day.

In principle, there are 5 methods to deal with non-wear to achieve a total daily activity level:

- Ignore it. Simply ignore the non-wear and average the total daily activity for the respective days. This will underestimate the daily activity level but will be easy to implement and close to real in cases with limited non-wear.
- Extrapolate the activity pattern from the other part of the day into the non-wear periods.
- Extrapolate the activity pattern from other similar days (time specific imputation). If sufficient days have been recorded the activity level from the same time of a similar day can be used as imputation.
- Only accept days having very high wear time i.e. >22 hours. All other days are excluded from the analysis.
- Report the data as percentage of wear time. This representation needs additional information about wear time to document that non-wear was not too high.

Most studies have used a minimum of 10 hours of wear during waking hours as cut off to be considered representative and count as a day (19). Assuming an average 8 hours of sleep this would mean having between 0 and 6 hours of “black box” activity on any given day. This will underestimate the daily physical activity depending on the activity level at the time of the day where the non-wear occurred.

Currently, there is no industry standard as to what amount of non-wear is accepted. The challenge is to define when to use which of the above approaches, i.e. how much non-wear is too much and when is it ok to ignore non-wear?

Handling sleep is also necessary and usually comes hierarchically after handling non-wear. Sleep is used to select part of the day where the patient is assumed to be awake. Sleep can be included in the analysis if it is important to evaluate the physical activity pattern during the entire day as opposed to only waking hours physical activity. The challenge of including sleep in the analysis is that it resembles sedentary activity and including or excluding sleep will have a substantial impact on estimates of sedentary time which is often an important analysis parameter. Therefore, it is common to exclude sleep when generating physical activity measures either by requiring participants to remove the device at bed time and resume wearing it upon waking or by subsequently handling it in data. The drawback of asking subjects to take the device off during sleep is that they sometimes forget to put it back on in the morning which can lead to substantial periods of non-wear. Handling sleep can otherwise be evaluated in two ways. Individualised time filters where a sleep diary or a sleep algorithm (20, 21) detects sleep or standard time filters where sleep is standardized each day at for instance 9 pm-6 am or 12 am-6 am. As expected, PA estimates vary considerably depending on the method used. A notable exception is moderate-to-vigorous PA (MVPA) where the estimates don't seem to change (22). Currently, there is not consensus as to how sleep, sedentary time and non-wear are differentiated including the correct classification of awake sedentary time and movement during sleep.

The challenge is to define how to handle sleep in the analysis.

SELECT THE WEAR PERIOD

For the data collected to be representative of the “real” physical activity level a certain amount of days must be recorded. This is to remove the impact of weekly recurring activities that impact the physical activity level. This could be weekends, once weekly soccer trainings or rare clinical events expected to alter the activity level in a population. As an example, the reproducibility of objectively measured daily MVPA in children (7–11 yr) using waist-mounted activity trackers required eight valid measurement days (10.1 hours of wear time) for boys and 10 days for girls (4). The same reliability is obtained with adults and the hip wear location requiring 6–9 days of measurement and 10–12 hours of wear time (23, 24). However, these numbers are taken from healthy individuals in a setting with close contact to the subjects and the device worn on the hip. In a patient population with limited contact and a larger number of clinical events in a wrist worn protocol it might be relevant to collect data for a longer period. If for example the total wear period is one week and by coincidence a clinical event happens causing pain; the physical activity level will be substantially lower and might no longer be representative of the normal level.

There is no clear definition of what sufficient wear time for a given protocol is, but it will depend on the activity tracker anatomical placement, how well the algorithm performs on the selected population and finally which endpoint is targeted. For instance, steps are particularly difficult to measure on a wrist worn device whereas it is simple on an ankle worn device simply because most step algorithms use the heel strike as indicator of a step. If it is expected that walking aids are used regularly, this will have a negative impact on the ability of the algorithm to detect physical activity as the movement pattern will be very different. Another concern is if the subjects are likely to participate

regularly in low-impact activities such as swimming, cycling or rowing which are not measured well using accelerometers.

The challenge is to define what a representative data collection period will be.

LACK OF AVAILABLE FIT-FOR-PURPOSE ALGORITHMS

Algorithms processing raw accelerometry data into meaningful endpoints should be fit-for-purpose. Fit-for-purpose in this context means that for physiological reasons algorithms are validated specifically for a certain age, disease, anatomical location and population. Validating an algorithm is the process by which we ensure that the measurement tool measures what it claims to measure. From a practical point of view, it can be difficult to check if a person was or was not sedentary at a certain point in time. For this purpose, several methods have been applied to be able to fact-check for instance the movement pattern or PA intensity of a person. This could be with the use of cameras or pressure sensitive insoles in the shoes. Unfortunately, there is a lack of validated, non-proprietary algorithms currently available for activity recognition. As an example, a study could include healthy subjects, with an age range of 12-70 years with age-specific algorithms automatically detecting non-wear (20, 21) and sleep (25, 26) in a wrist-worn protocol focusing on cut points (3, 27) as an endpoint. Figure 2 below, shows examples of algorithms fit-for-purpose but it also highlights that they have not been validated in all age groups.

The question is whether the data output would be considered valid in a submission?

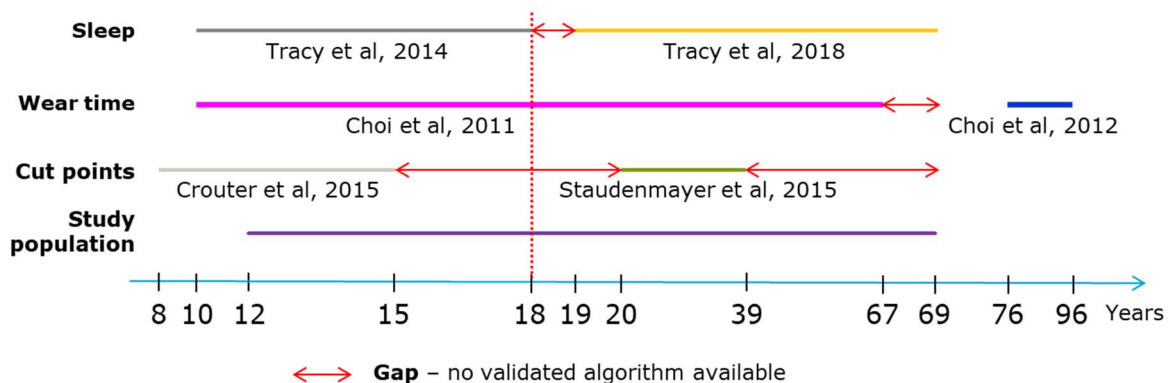


Figure 2. The figure shows age gaps in the selected algorithms relative to the study population

Another aspect in the fit-for-purpose discussion is the quality of algorithms that are validated, i.e. are they good enough even though they are disease specific and validated in the appropriate age, population and anatomical location? An example could be step detection algorithms. Steps has been used extensively in publications and achieving >10000 steps per day has been advocated by the Danish Health Authorities. At first sight, steps are easy and intuitive to understand and suitable for adoption as an endpoint in clinical trials. In contrast to the other endpoints based on metabolic cost it describes movement without acceleration and is determined based on thresholds i.e. either a step was detected, or no step was detected. The step length or energy going into the step is therefore not captured. The device anatomical location will have a large impact on the validity of a step algorithm to detect a step. Most step algorithms use the impact made by the heel strike to detect that a step has been taken. Thus, placement of the device on the ankle will create a clear spike in the signal and allow an accurate prediction of a step. However, the most common anatomical placement of a wearable device is the wrist and although some of the newer algorithms have attempted to use the arm swing to detect a step, the heel strike signal is still the most commonly used and the attenuated signal on the wrist can lead to false negative results (28).

The challenge is that several algorithms exist that detects steps, however arguably no algorithm currently exists with the ability to detect steps accurately in a free-living environment even though they are “validated”.



THE CHALLENGE OF HANDLING THE SHIFT IN AGE FROM ADOLESCENT TO ADULT

Algorithms are very often specific to either children, adolescents or adults. The reason is that the physical activity pattern and physical activity guidelines for children is different from adults. Children tend to move in high intensity bursts followed by short breaks whereas adults move more homogenously. Also, factors such as leg length (which is associated with age) cause changes in step detection. Because several of the algorithms are age-specific there is a challenge in cases where a trial lasts several years, and subjects shift from adolescents to adults during study conduct. Which algorithm should be chosen? An option would be to apply an adult algorithm for subjects aged 15 years and above at time of screening and apply an adolescent algorithm for subjects up to and including the age of 14 years at screening for a trial lasting 4 years. However, this could create a scenario where a subject enrolled at age 15 and terminating the study early before becoming an adult (18 y) would have data based on an adult algorithm.

The challenge is how to handle changes in age during trials.

COMPARING RESULTS BETWEEN OR WITHIN TRIALS IS DIFFICULT

Being able to compare results between or within trials can have substantial value. However, the lack of standardization within the area of physical activity data makes it very difficult. The major reason for the lack of standardization is the immaturity of the available algorithms used in the area and this must be taken into account when attempting to compare two sets of results.

As an example, physical activity algorithms have only recently started to use raw data as input. Earlier it was common for algorithms to use so-called counts as input to create other endpoints. Counts are processed from raw data and provides a meaningful measure of PA in itself. This means that it will be challenging to disentangle how algorithms derived from raw data has handled for instance external noise compared to an algorithm based on counts and what impact it has on the results.

Another example is cut points, which provide information about time spent in different activity levels i.e. sedentary, light, moderate or vigorous activity in minutes per day. If the algorithms don't use the same cut point categories a comparison can be hard. For instance, the cut-points used in a study by Gonzales et al. was 3600 and 6199 counts/min for moderate and vigorous activity respectively, whereas it was 7320 and 21708 counts/min for the algorithm used by Crouter (27, 29). Finally, the thresholds defining the intensity in each cut point also differs, as well as the definition of a count, the sampling frequency and the difference in target MET score for the algorithms.

Even within a trial a direct comparison can be difficult. For instance, including multiple populations in a trial would require multiple algorithms, unless the algorithm was validated in all the populations. The same applies to subjects with different ages where age specific algorithms are needed. Because algorithms can be very different a direct comparison between children and adults can be meaningless.

SECOND PART

NOVEL ENDPOINTS AND THE IMPACT ON DATA GRANULARITY AND DATA VOLUME

Traditional endpoints such as steps can be derived from data aggregated into day-level, i.e. steps/day. This will keep the data volume at a manageable level. Additional records would be present to quantify the number of hours of sleep and non-wear. However, these numbers will not provide information about when non-wear or sleep happened and getting to know patterns in data or evaluating or comparing the performance of the algorithms will be harder. For instance, are sleep periods fragmented or of longer duration than expected; do the sleep periods often appear during the day for no reason, is vigorous activity detected during night time and if yes, when.

Furthermore, novel endpoints based on pattern recognition would be derived from minute level or even raw data. Examples could be acute one-day studies or an endpoint where the physical activity or movement pattern surrounding a clinical event is of interest. Also, two subjects might have the same average daily amount of activity, but because of a disease of interest the activity pattern during the day might look different between the two. This would be possible to visualize in a higher granularity dataset.

DATA STRUCTURE

A common understanding of SDTM IG 3.2 is that all collected data must be tabulated in SDTM format. When an endpoint is based directly on minute-level data, the relative source records used in the analysis must be submitted and included in the SDTM submission package.

This chapter deals with the challenge of adding minute level data into SDTM for submission purposes. In this example, the collection of 25 derived parameters, performed every minute for 10 weeks with just a small sample of participants (300), would lead to a breath-taking 800 million records in the maximum scenario. It quickly becomes clear that this data volume in SDTM could weigh down internal systems and solutions to reduce the amount of data are needed.

In our case we have selected a vendor who is able to deliver the required data in SDTM format. Given the amount of data planned to be received from the vendor, we performed an analysis and impact assessment of the data structure. To reduce the amount of data to be transferred at each load and still create an SDTM compliant package, we restricted the variables we received to the bare minimum and handled the remaining variables internally. This was for instance the ones where the value is identical in the entire dataset i.e. STUDYID, the ones with one-to-one match: --TEST and --TESTCD, and the ones requiring additional information not available to the provider like --DY.

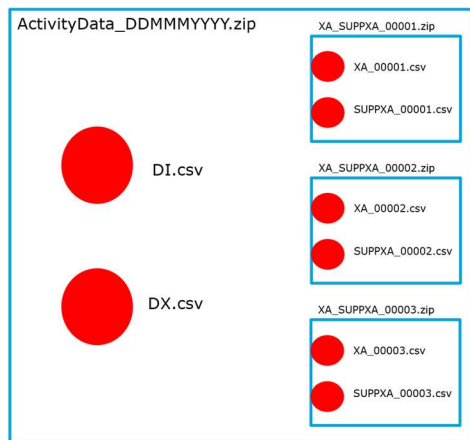


Figure 3. The figure shows the structure of the zip file and relative content received at each data transfer, blue squares represent zip files while red circles represent csv files.

We decided to receive the data from the vendor in multiple files in Comma Separated Values format (.csv). Due to the planned amount of data, for XA and SUPPXA domains, we received the .csv files per subject (Fig 3). This serves the purpose of dealing with smaller files and, at the same time, we remain consistent with the source data while splitting the domain in the creation of .xpt files for submission.

When dealing with external providers and new types of data, we recommend being clear and exhaustive on documentation and data specifications as data handling in-house will be much easier.

SELECTING THE SDTM DOMAINS

We decided to collect the actual physical activity data in a sponsor defined findings domain, named XA (Physical Activities), already consolidated in the organization. This domain is linked to its supplemental qualifier domain SUPPXA, where we decided to collect additional information about the time zone to be able to work with data without breaks associated with daylight savings. We used IDVAR as the XAGRPID variable from the parent domain, grouping all the 25 derived parameters for a specific timepoint of a specific subject. This was done to avoid having the time zone information replicated for each of the 25 parameters per minute. In addition, we received the results in standard units from the vendor, to avoid redundant derivations.

XA and SUPPXA contain the subject related information, while the device related information is captured in the DI domain (Study Device Identifiers) and DX domain (Device Exposure) according to SDTMIG-MD. DI is a special-purpose domain designed for the submission of information that identifies a specific device. Here we included 5 parameters which are discussed in the next section.

DX is an interventions domain that records the details of a subject's exposure to a medical device under study. With

this domain we can link the devices with SPDEVID (Sponsor Device Identifier) to each single subject with USUBJID (Unique Subject Identifier).

XA is linked with SUPPXA with the combination of USUBJID and XAGRPID as IDVARVAL. The value of IDVAR is defaulted internally as “XAGRPID” (red boxes in Figure 4). XA and DX domains are linked with a combination of USUBJID and the datetime of collection for each finding, that needs to fall within a timeframe defined by DXSTDTC and DXENDTC. This also identifies the device used for collecting the data in the XA domain (green boxes in Figure 4). The SPDEVID variable is linking DX and DI domains, providing more information about the device used (yellow boxes in Figure 4). All the datetime information in parental domains are collected using the local time, therefore XA and DX domains are linked directly with XADTC and DXSTDTC/DXENDTC without using the UTC datetime in SUPPXA. In general, in device trials, the same device could possibly be used by multiple subjects or multiple devices could possibly be used by the same subject. For instance, in the first case the same SPDEVID would be repeated in different records of the DX domain. The bundle of domains containing device information and findings are linked together as in the image below.

USUBJID	XAGRPID	XATESTCD	XAORRES	XAORRESU	XADTC	
NNXXXX-YYYY/123456	86	AC11WEAR	Y		2019-07-25T14:12	XA
NNXXXX-YYYY/123456	86	AT14SLA	N		2019-07-25T14:12	
NNXXXX-YYYY/123456	86	AT18SLAD	N		2019-07-25T14:12	
NNXXXX-YYYY/123456	86	S15ADLI	30	sec	2019-07-25T14:12	
NNXXXX-YYYY/123456	86	S15ADMO	30	sec	2019-07-25T14:12	

USUBJID	SPEVID	DXLAT	DXSTDTC	DXENDTC	
NNXXXX-YYYY/123456	XYZ1A37180280	RIGHT	2019-05-28T16:22	2019-09-24T10:11	DX

USUBJID	IDVARVAL	QNAM	QVAL	SPDEVID	DIPARMCD	DIVAL	
NNXXXX-YYYY/123456	86	UTCDATTM	2019-02-28T21:22Z	XYZ1A37180280	TYPE	Kinesiology Ambulatory Recorder	DI
NNXXXX-YYYY/123456	86	TIMEZONE	America/New_York	XYZ1A37180280	MANUF	Provider Name	
				XYZ1A37180280	MODEL	XYZ01	
				XYZ1A37180280	SERIAL	XYZ1A37180280	
				XYZ1A37180280	FDAUDI	FDAUDI Number	

Figure 4. The figure shows a sample of the collected data, mapped in the selected domains, and their connections.

DR domain (Device-Subject Relationships) is a special-purpose domain that links each subject to the devices they may have been exposed to. We decided to derive this domain from the DX domain. It provides a single, consistent location to find the relationship between a subject and a device, regardless of the device or the domain in which subject-related data may have been collected or submitted. This can be valuable as one subject may have more than one device combination.

SPLIT DOMAINS

According to FDA Study Data Technical Conformance Guide, each dataset should be provided in a single transport file and datasets greater than 5 gigabytes (GB) in size should be split into smaller datasets no larger than 5 GB. A first estimation on a complete package of XA and SUPPXA domains, in the unrealistic scenario of complete wear and fully compliant SDTM package, lead to a complete data dimension of about 200GB.

The suggested approach on the same document of splitting by categories and subcategories, in our case study does not suffice the requirement on the maximum dimension of each single dataset.

The only relevant variable present in the XA domain, which would allow to create split datasets compliant with this requirement is USUBJID. Therefore, we decided to have one XA split domain for each participant in the trial, following the suggested naming convention, concatenating the domain name with a running number, from XA1 to XA300.

However, supplemental qualifier domains must be split following the principle of their parental domain, therefore SUPPXA must be divided in 300 components named SUPPXA1 to SUPPXA300.

But since the dataset deliverables are supposed to be in xpt format, this is not possible, as the SAS procedure to create XPT files is limited to 8 characters for the file name. For subjects flagged with the third digit number (from SUPPXA100 to SUPPXA300) this limit is exceeded. A consistent approach would be the usage of the SQ suffix instead of SUPP, following the approach of AP domains.

Splitting the domains due to the data dimension is a suitable approach to handle data transfer, however it seems that the naming convention doesn't take this number of splits into consideration. On a submission point of view, would another approach be considered valid by the authorities?

PARAMETERS COLLECTED

In the XA domain the parameters collected have 4 levels of detail, the measurement, the algorithm used, the relative age and the intensity of activity, if relevant.

Number of TESTCD's	ALGORITHM	ENDPOINT	ADOLESCENTS	ADULTS	SEDENTARY	LIGHT	MODERATE	VIGOROUS
1	CHOI 2011	WEAR						
4	AXIS COUNTS	ACTIVITY						
2	TRACY 2014/2018	SLEEP						
2	UWF STEPS	ACTIVITY						
6	STAUDENMAYER 2015	ACTIVITY						
5	VA CROUTER 2015	ACTIVITY						
5	VM CROUTER 2015	ACTIVITY						

Figure 5. The figure shows the categorisation of the parameters collected.

Given the file size dimension, we found it important to work on a proper categorisation of the parameters collected in the XA domain, to allow an easy and consistent subset of the relevant findings across the records. We decided to derive a 2-level categorisation, where XACAT represents the type of endpoint, XASCAT contains both the algorithm used and the age classification. The intensity of activity is detailed in XATESTCD. This multiple level categorisation enables an easy and direct subset of the necessary information for each endpoint, with each algorithm.

The wear algorithm has Boolean results and it is collected as a flag (Y or N), it is independent from the age of the subject and guides the decision about the relevance to collect all the other parameters, i.e. if wear is flagged as Y then the remaining 24 parameters are received. The inclusion of this flag as decision maker for the minute level data led to a substantial reduction of the number of records collected. The choice of the XATESTCD follows the same categorisation.

Figure 7 shows a sample of the DIPARMCD values in the DI Domain. The SDTMIG-MD specifies that these are required for all device submissions where any device-specific information is included and is encouraged for the remainder. Our aim is to create a full SDTM compliant data package, regardless of the type of submission, therefore we also included the FDAUDI (the FDA's UDI identifier) parameter as required, even if it is only for post-marketing studies.

XATESTCD	XATEST	XACAT	XASCAT
AC11WEAR	CHOI 2011 WEAR STATUS	WEAR	CHOI 2011 ALL AGES
AT14SLA	TRACY 2014 SLEEP STATUS ADOLESCENT	SLEEP	TRACY 2014 ADOLESCENT
AT18SLAD	TRACY 2018 SLEEP STATUS ADULT	SLEEP	TRACY 2018 ADULT
S15ADLI	STAUDENMAYER 2015 ADULT LIGHT	CUT POINTS	STAUDENMAYER 2015 ADULT
S15ADMO	STAUDENMAYER 2015 ADULT MODERATE	CUT POINTS	STAUDENMAYER 2015 ADULT
S15ADVI	STAUDENMAYER 2015 ADULT VIGOROUS	CUT POINTS	STAUDENMAYER 2015 ADULT
S15ADVM	STAUDENMAYER 2015 ADULT MOD AND VIG	CUT POINTS	STAUDENMAYER 2015 ADULT
S15ADX	STAUDENMAYER 2015 ADULT SEDENTARY TIME	CUT POINTS	STAUDENMAYER SED 2015 ADULT
S15ADXNS	STAUDENMAYER 2015 ADULT NON SEDENTARY TIME	CUT POINTS	STAUDENMAYER SED 2015 ADULT
UF19STA	UWF 2019 STEPS ADOLESCENT	STEPS	UWF 2019 ADOLESCENT
UF19STAD	UWF 2019 STEPS ADULT	STEPS	UWF 2019 ADULT
VAC15AES	VA CROUTER 2015 ADOLESCENT SEDENTARY	CUT POINTS	VA CROUTER 2015 ADOLESCENT
VAC15ALI	VA CROUTER 2015 ADOLESCENT LIGHT	CUT POINTS	VA CROUTER 2015 ADOLESCENT
VAC15AMO	VA CROUTER 2015 ADOLESCENT MODERATE	CUT POINTS	VA CROUTER 2015 ADOLESCENT
VAC15AVI	VA CROUTER 2015 ADOLESCENT VIGOROUS	CUT POINTS	VA CROUTER 2015 ADOLESCENT
VAC15AVM	VA CROUTER 2015 ADOLESCENT MOD AND VIG	CUT POINTS	VA CROUTER 2015 ADOLESCENT
VMC15AES	VM CROUTER 2015 ADOLESCENT SEDENTARY	CUT POINTS	VM CROUTER 2015 ADOLESCENT
VMC15ALI	VM CROUTER 2015 ADOLESCENT LIGHT	CUT POINTS	VM CROUTER 2015 ADOLESCENT
VMC15AMO	VM CROUTER 2015 ADOLESCENT MODERATE	CUT POINTS	VM CROUTER 2015 ADOLESCENT
VMC15AVI	VM CROUTER 2015 ADOLESCENT VIGOROUS	CUT POINTS	VM CROUTER 2015 ADOLESCENT
VMC15AVM	VM CROUTER 2015 ADOLESCENT MOD AND VIG	CUT POINTS	VM CROUTER 2015 ADOLESCENT
XAXIS	AXIS X COUNTS	COUNTS	ALL AGES
YAXIS	AXIS Y COUNTS	COUNTS	ALL AGES
ZAXIS	AXIS Z COUNTS	COUNTS	ALL AGES
ZVECTMG	VECTOR MAGNITUDE COUNTS	COUNTS	ALL AGES

Figure 6. The figure shows the combination of XATESTCD, XATEST, XACAT and XASCAT selected.

SPDEVID	DIPARMCD	DIVAL
XYZ1A37180280	TYPE	Kinesiology Ambulatory Recorder
XYZ1A37180280	MANUF	Provider Name
XYZ1A37180280	MODEL	XYZ01
XYZ1A37180280	SERIAL	XYZ1A37180280
XYZ1A37180280	FDAUDI	FDAUDI Number

Figure 7. The figure shows a sample of the parameters collected for each device in the DI domain.

REQUIRED VARIABLES

SDTM compliance requires a lot of derivations and addition of replicated values for each variable in each domain. This approach serves the purpose for standard clinical trials data and allows for an accurate review as each record describes itself. However, in very large datasets it becomes cumbersome to repeat the same value for each of the millions of records included in domains containing device data.

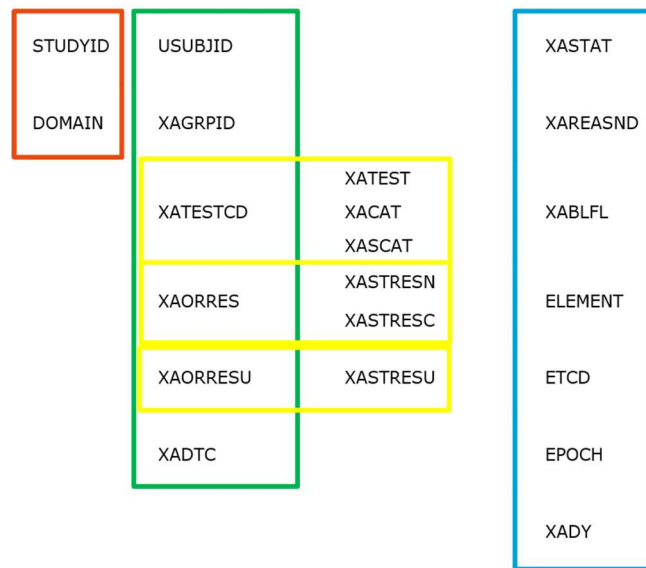


Figure 8. The figure shows a categorisation of the source variables against the supplemental information required in SDTM.

In Figure 8, the variables in the red box are simple replications of the same value across the complete domain. In this case study, the addition of the variable STUDYID alone is leading to a projection in the worst-case scenario of potentially more than 8gb in compressed SAS datasets. The variable STUDYID alone, would exceed the maximum data dimension allowed for XPT files.

The variables in the green box are the only one with central importance. These variables contain the values we are receiving from the vendor. However, given their inclusion in the SDTM domain, following the general principles of the SDTM-IG 3.2 for findings domain, the inclusion of some more variables is mandatory.

In fact, in the yellow boxes the variables populated as one-to-one derivation from the source variables are included. XATEST, XACAT, XASCAT can simply be derived from the XATESTCD variable with no exceptions. XAORRES implies the derivation of XASTRESN and XASTRESC, while XAORRESU the derivation of XASTRESU. Since we are receiving the findings already in standard units from the vendor, these derivations are simple copies (or for XASTRESN character to numeric conversions) of their relative source variable in the green box.

In the blue box the variables are usually derived in findings domains. The derivation of XABLFL doesn't bring much value in the domain, since the complete first minute (XAGRPID=1) would be flagged as baseline value, however in reality it could not be considered a baseline assessment, as in the big data perspective the single minute doesn't have much value, it's the aggregated results that would be considered relevant, finding patterns in the big picture. ELEMENT and ETCD are required variables, while EPOCH is required from FDA. But the complete data collection of device data in this trial belongs to the same ELEMENT and EPOCH, therefore these variables, would just be derived as a simple replication of the same value across the whole domain, like STUDYID and DOMAIN.

Dealing with massive amounts of data is challenging both for pharmaceutical companies and authorities. Are these derivations to follow the SDTM specifications really needed? Would it be better to provide duplicate information in different variables to reduce the data amount to the minimum necessary? Is the effort of ensuring derivation of domains and qualifiers and mapping to prespecified SDTM formats required when dealing with unusual and exploratory data?

CONCLUSIONS

Wearable medical devices can objectively measure outcomes in a real life setting and generate new types of datasets supporting new types of endpoints. However, with new types of data sets and endpoints comes new challenges and decisions to be made. A successful trial will need to carefully select the data collection setup and consider the best way of handling the data from endpoint design to submission to increase the likelihood for success. In this context, more work needs to be done to understand and validate potential applications for patient-generated health data (i.e., from wearable devices) in collecting RWE. In this paper we have discussed some of the challenges we have seen while preparing our first generation of SDTM data based on an FDA class II wrist worn activity tracker. We hope that this will create awareness of the potential in wearable data and help kickstart the discussion about ways to improve data quality and how traditional SDTM data format will become fit-for-purpose for the new data types coming.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Author Name: Martin Gram
Company: Novo Nordisk A/S
Address: Vandtårnsvej 108-110
City / Postcode: 2860 Søborg
Email: xmgm@novonordisk.com
Web: www.novonordisk.com

Author Name: Gianluca Mortari
Company: Novo Nordisk A/S
Address: Vandtårnsvej 108-110
City / Postcode: 2860 Søborg
Email: glcm@novonordisk.com
Web: www.novonordisk.com

Brand and product names are trademarks of their respective companies.

REFERENCES

1. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and science in sports and exercise*. 1998;30(5):777-81.
2. Freedson PS, Sirard J, Debold E, Pate R, Dowda M, Trost S, et al., editors. CALIBRATION OF THE COMPUTER SCIENCE AND APPLICATIONS, INC. (CSA) ACCELEROMETER 2561997 May 1997.
3. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *Journal of applied physiology*. 2015;119(4):396-403.
4. Byrom B, Rowe DA. Measuring free-living physical activity in COPD patients: Deriving methodology standards for clinical trials through a review of research studies. *Contemp Clin Trials*. 2016;47:172-84.
5. Leinonen AM, Ahola R, Kulmala J, Hakonen H, Vaha-Ypya H, Herzig KH, et al. Measuring Physical Activity in Free-Living Conditions-Comparison of Three Accelerometry-Based Methods. *Frontiers in physiology*. 2016;7:681.
6. Fairclough SJ, Noonan R, Rowlands AV, Van Hees V, Knowles Z, Boddy LM. Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers. *Med Sci Sports Exerc*. 2016;48(2):245-53.
7. Rowlands AV, Rennie K, Kozarski R, Stanley RM, Eston RG, Parfitt GC, et al. Children's physical activity assessed with wrist- and hip-worn accelerometers. *Med Sci Sports Exerc*. 2014;46(12):2308-16.
8. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181-8.
9. Doherty A, Jackson D, Hammerla N, Plotz T, Olivier P, Granat MH, et al. Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PLoS One*. 2017;12(2):e0169649.
10. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol Meas*. 2014;35(11):2191-203.
11. Ellis K, Kerr J, Godbole S, Staudenmayer J, Lanckriet G. Hip and Wrist Accelerometer Algorithms for Free-Living Behavior Classification. *Med Sci Sports Exerc*. 2016;48(5):933-40.
12. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. *Med Sci Sports Exerc*. 2013;45(5):964-75.
13. Kim Y, Lee JM, Peters BP, Gaesser GA, Welk GJ. Examination of different accelerometer cut-points for assessing sedentary behaviors in children. *PLoS One*. 2014;9(4):e90630.
14. Tudor-Locke C, Barreira TV, Schuna JM, Jr., Mire EF, Chaput JP, Fogelholm M, et al. Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE). *The international journal of behavioral nutrition and physical activity*. 2015;12:11.
15. Schneller MB, Bentsen P, Nielsen G, Brond JC, Ried-Larsen M, Mygind E, et al. Measuring Children's Physical Activity: Compliance Using Skin-Taped Accelerometers. *Med Sci Sports Exerc*. 2017;49(6):1261-9.
16. Colley R, Connor Gorber S, Tremblay MS. Quality control and data reduction procedures for accelerometry-derived measures of physical activity. *Health reports*. 2010;21(1):63-9.
17. Herrmann SD, Barreira TV, Kang M, Ainsworth BE. Impact of accelerometer wear time on physical activity data: a NHANES semisimulation data approach. *British journal of sports medicine*. 2014;48(3):278-82.
18. Toftager M, Kristensen PL, Oliver M, Duncan S, Christiansen LB, Boyle E, et al. Accelerometer data reduction in adolescents: effects on sample retention and bias. *The international journal of behavioral nutrition and physical activity*. 2013;10:140.
19. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nystrom C, Mora-Gonzalez J, Lof M, et al. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports medicine*. 2017;47(9):1821-45.
20. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc*. 2011;43(2):357-64.
21. Choi L, Ward SC, Schnelle JF, Buchowski MS. Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. *Med Sci Sports Exerc*. 2012;44(10):2009-16.

22. Meredith-Jones K, Williams S, Galland B, Kennedy G, Taylor R. 24 h Accelerometry: impact of sleep-screening methods on estimates of sedentary behaviour and physical activity while awake. *J Sports Sci.* 2016;34(7):679-85.
23. Aadland E, Ylvisaker E. Reliability of Objectively Measured Sedentary Time and Physical Activity in Adults. *PloS one.* 2015;10(7):e0133296.
24. Aadland E, Ylvisaker E. Reliability of the Actigraph GT3X+ Accelerometer in Adults under Free-Living Conditions. *PloS one.* 2015;10(8):e0134606.
25. Tracy DJ, Xu Z, Choi L, Acra S, Chen KY, Buchowski MS. Separating bedtime rest from activity using waist or wrist-worn accelerometers in youth. *PLoS One.* 2014;9(4):e92512.
26. Tracy JD, Acra S, Chen KY, Buchowski MS. Identifying bedrest using 24-h waist or wrist accelerometry in adults. *PLoS One.* 2018;13(3):e0194461.
27. Crouter SE, Flynn JJ, Bassett DR, Jr. Estimating physical activity in youth using a wrist accelerometer. *Med Sci Sports Exerc.* 2015;47(5):944-51.
28. Hickey A, John D, Sasaki JE, Mavilia M, Freedson P. Validity of Activity Monitor Step Detection Is Related to Movement Patterns. *J Phys Act Health.* 2016;13(2):145-53.
29. Gonzalez LM, Peiro-Velert C, Devis-Devis J, Valencia-Peris A, Perez-Gimeno E, Perez-Alenda S, et al. Comparison of physical activity and sedentary behaviours between young haemophilia A patients and healthy adolescents. *Haemophilia.* 2011;17(4):676-82.