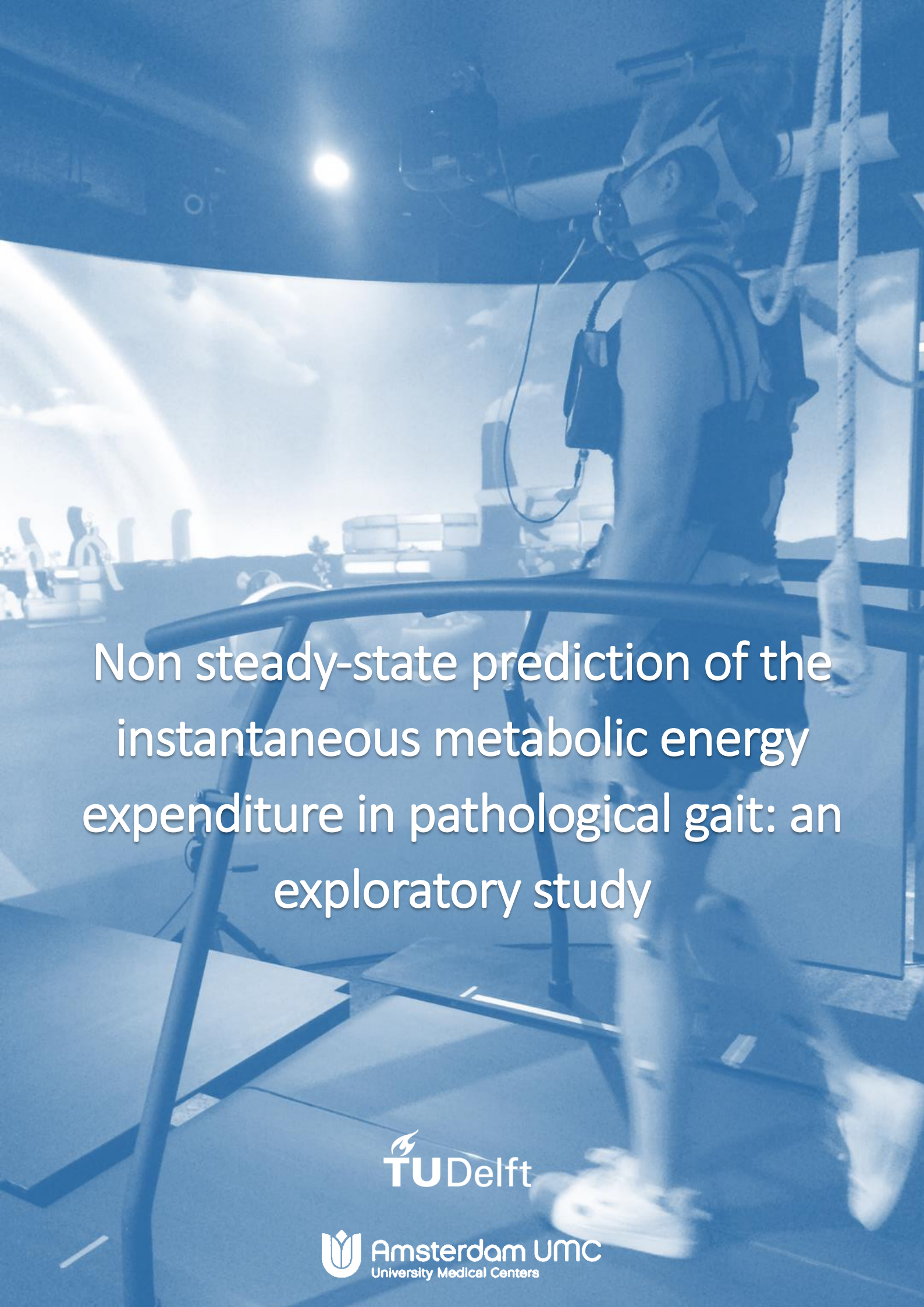


# Non steady-state prediction of the instantaneous metabolic energy expenditure in pathological gait: an exploratory study

TU Delft

Amsterdam UMC
University Medical Centers

Thesis manuscript

# Non steady-state prediction of the instantaneous metabolic energy expenditure in pathological gait: an exploratory study

by

Suzanne Joëlle Filius
4688244

Master Thesis Biomedical Engineering (BM51032)

Department of Biomedical Engineering
Faculty of Mechanical, Maritime and Material Engineering
Delft University of Technology
BioMechaMotionlab

To be defended on September 5th, 2019

Thesis Committee:

| | |
|---|---|
| Prof. dr. ir. J. Harlaar | TU Delft, supervisor |
| Prof.Dr.-Ing. H. Vallery | TU Delft |
| Dr. D. Dodou | TU Delft |
| Dr.ir. J.C.F. de Winter | TU Delft |

TU Delft

# Abstract

**BACKGROUND** • In a variety of patients with locomotive disorders gait efficiency is often assessed using mobile metabolic gas analysis during rehabilitative interventions. Conventionally, these measurements take 6 minutes per evaluated condition (e.g. with/without ankle-foot-orthosis; AFO). This duration is required since there is a dynamical delay between the instantaneous metabolic energy expenditure (ImEE), and the respiratory response measured at the mouth. Moreover, the breath-by-breath data is sparely sampled and noisy. Gait efficiency is therefore computed from an averaged (i.e. of 1-3 minutes) respiratory response during a steady-state metabolism (i.e. reached after 3-4 minutes). This is time consuming and can be exhausting for patients with severe gait impairments.

In up-coming *human-in-the-loop* techniques, fast (2-3 minutes) predictions of the ImEE are made using an Instantaneous Cost Mapping (ICM) model. The ICM is based on a first-order response with a known time constant ($\tau$) to a change in an external load (e.g. another walking speed). The $\tau$ is either identified on a subject-specific basis or set to a fixed average time constant already identified for healthy adults (42s; Selinger & Donelan, 2014).

It is unknown whether this technique could be applied in patient populations with additional gait impairments. When applicable, the ICM model could speed-up the assessment of the gait efficiency and open ways to make human-in-the-loop protocols feasible device optimization (e.g. AFO tuning) in rehabilitation.

**AIM** • This study investigates whether the subject-specific and/or general ICM model could reduce the required measurement duration in a variety of patients that cope with gait impairments as consequence of a neurology or neuromuscular disease. Secondarily, it explores whether there are differences in the identified subject-specific $\tau$ among the pathologies.

**METHODS** • Post hoc metabolic data, recorded in the period of 2006 to 2019 within the Amsterdam University Medical Centers (UMC) was collected, containing walking trials of patients with Multiple Sclerosis, Cerebral Palsy, several neuromuscular disorders, healthy adults and typically developing children. The ICM model was tested for the subject-specific and general $\tau$ to estimate the ImEE. First, using full measurement durations and later the identified reduced measurement durations. The model outcome was assessed for individually correctness.

**RESULTS** • Results show ($n$ = 28) that the ICM model correctly estimates the ImEE using the total conventional measurement duration (subject-specific: 96%; general: 89% individually correctness). For the identified reduced measure duration (subject-specific: 2m33s; general: 3m:04s) reduced the individually correctness to 57%. No differences among the groups were found for the subject-specific $\tau$ and the average ($\tau$ = 41.7s ± 13.5s) was similar to the reported $\tau$ for healthy adults.

**CONCLUSIONS** • Based on the results, it can be concluded that the ICM model show similar results to healthy adults from literature. This offers prospect for the clinical application of the model in rehabilitation depending on the purposes of use. Clinicians should consider the balance between the individually correctness and measurement duration of the predictions. Results should be treated with caution due to the small and heterogeneous sample size. Future research is needed to identify differences in subject-specific $\tau$ among pathologies.

**Keywords**: human-in-the-loop, instantaneous-cost-mapping, wearable device optimization, metabolic gas analysis, non-steady-state prediction, gait impairments.

# Content

# 1. Introduction

## *1.1. Human locomotion*

"Humans are skilled walkers" (Collins, Wiggin, & Sawicki, 2015, *p*.1). Between approximately the 10th to 18th month of life a human baby starts to make its first steps and learns how to walk. After a while, walking becomes a natural automated and unconscious movement coordination pattern that requires barely mental load. However, from a biomechanical perspective bipedal gait is a "phenomenon of the most extraordinary complexity" (Saunders, Inman, & Eberhart, 1953, *p*.543). It involves smooth displacement of the body's center of mass (COM) through space in a forward direction, transported by the inverted pendulum like motion of the legs (Waters & Mulroy, 1999; Donelan, Kram, & Kuo, 2002). In the step-to-step transition, positive and negative mechanical work is performed to accomplish the redirection of the COM (Kuo, Donelan, & Ruina, 2005). In humans, positive, negative and even no mechanical work can cost metabolic energy. Due to the concentric, eccentric and isometric contractions of the muscles (Collins et al., 2015). Although human locomotion is highly optimized, walking still costs a significant amount of metabolic energy due to these muscle contractions (Waters & Mulroy, 1999; Kuo et al., 2005; Collins et al., 2015). Humans naturally tent to keep this metabolic energy expenditure (mEE) low by optimizing different determinants of gait in such way that less energy is dissipated, and more energy can be conserved (Saunders et al., 1953; Waters & Mulroy, 1999; Collins et al., 2015). Healthy humans seem to walk with a particular combination of step length, frequency and width that seems to be energetically optimal (Kuo, 2001; Bertram, & Ruina, 2001; Donelan, Kram, & Kuo, 2001; Kuo et al., 2005; Collins et al., 2015). Only a small adaptation in gait pattern, trunk, or upper limb motion can result in a change in mEE during walking (Waters & Mulroy, 1999).

## *1.2. Metabolic energy expenditure*

The skeletal muscles in the human body use chemical energy to generate the required mechanical work (Collins et al., 2015). This chemical energy is also called adenosine triphosphate (ATP) (Koller, Gates, Ferris, & Remy, 2017). During muscle contraction, the actin and myosin filaments of the muscle fibers attach and form cross-bridges, causing the sarcomeres to become smaller. The actin and myosin are continuously attaching and detaching, whereby detaching costs chemical energy in the form of ATP. ATP is replenished by creatine phosphate, and creatine phosphate is replenished in the mitochondria. During this process, oxygen and glucose are consumed and carbon dioxide and water are produced (Selinger & Donelan, 2014). The supply and disposal of the oxygen and carbon dioxide is made possible by the respiratory and cardiovascular system of the body. The oxygen and carbon dioxide are transferred by the red blood cells in the blood towards the lungs, where respiratory gas exchange take place. The gases are in- and exhaled through the mouth. The oxygen and carbon dioxide in the breathing air can give an indication of how much chemical energy the body has produced from the inhaled oxygen. This can be used to estimate the mEE of walking (Selinger & Donelan, 2014).

## *1.3. Pathological gait*

Neurology and neuromuscular diseases such as multiple sclerosis (MS), cerebral palsy (CP), and post-polio syndrome (PPS) are often accompanied by gait disabilities. In response to these gait disabilities, patients tent to adapt their gait pattern with compensatory strategies. Compensative strategies can cause redistributions in the power supply from, for example, highly efficient ankle musculotendons to less efficient hip musculotendons. This can result in an increased mEE during gait (Saunders et al., 1953; Zamparo, Franscescato, de Luca, Lovati, & Prampero, 1995; Waters & Mulroy, 1999; Bregman, Harlaar, Meskers, & de Groot, 2012). Moreover, elevated muscle tone often seen in CP patients, arising from co-contraction or spasticity, can elevate the mEE even more (Unnithan, Dowling, Frost, & Bar-Or, 1996; Schwartz, 2007). The magnitude of the associated increase in mEE depends on "the severity of

the disability and the patients' cardiovascular and musculoskeletal fitness" (Waters & Mulroy, 1999, *p*.208).

Orthotic surgery, gait retraining and assistive wearable devices can reduce the elevated mEE in pathological gait and helps the patient to participate in more activities of daily living (Slade, Troutman, Kochenderfer, Collins, & Delp, 2019). One of the most important objectives of rehabilitation. Orthopedic devices such as ankle-foot-orthoses (AFO) are frequently prescribed within the aforementioned neurologic and neuromuscular diseases. Important for such devices is that the orthoses are correctly aligned and adapted to the needs of the patient and do not result in deteriorations in mEE. To evaluate whether the rehabilitation intervention was effective and/or if the gait efficiency of the patient improves over time, the mEE during gait is commonly assessed at multiple occasions during rehabilitation services (Schwartz, 2007; Slade et al., 2019; Waterval, Nollet, Harlaar, & Brehm, 2017).

### *1.4. Indirect calorimetry*

To measure the gait efficiency, an *indirect calorimeter* is used, measuring the breathing composition during a *metabolic gas analysis*. During the gas analysis, a mask is placed over the nose and mouth so that the in- and exhaled gases flow through a volume turbine, which measures the respiratory flow. An additional gas sample line measures the composition of the breathing gas (Brehm, Harlaar, & Groepenhof, 2004), see Figure 1. At every breath, a sample is taken with the volumes of oxygen and carbon dioxide (Felt et al., 2015). This information is sent to the corresponding indirect colorimeter and the results are sent forward to a computer. These results are used to compute an estimation of the mEE. However, the preparation, measurement and data processing of this technique deserve caution.

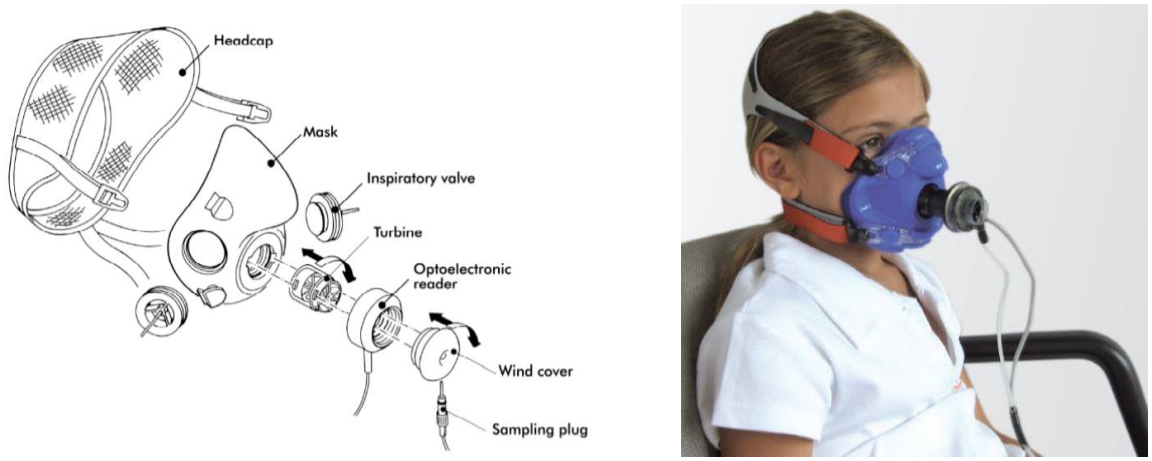



**Figure 1**. Illustration the indirect calorimeter mask, gas sample line, and volume turbine. *Note*: retrieved from http://www.cosmed.com/

#### 1.4.1. Preparation and measurement

It is important that participants do not eat prior testing to reduce the variability in mEE due to digestion. The time for which patients are asked to fast differs from 1.5 up to 12 hours prior testing (Brehm, Becher, & Harlaar, 2007; Brehm, Knol, & Harlaar, 2008; Quesada, Caputo, & Collins, 2016). Moreover, the indirect calorimeter needs to be calibrated within the measurement environment. If the equipment is not correctly calibrated, the measurements could contain inaccuracies (Brehm et al., 2004).

The gas analysis often starts with a measurement of the rest metabolism which represents the internal basal metabolism. During this measurement, the participant is asked to sit, stand or lay down as still as possible for a period of 5 to 10 minutes (Brehm et al., 2007; Schwartz, 2007). This rest measurement is often followed by a 6- or 12-minute walk test on a treadmill or an over-ground marked

track to assess the gait efficiency of the participant on a comfortable self-selected or pre-defined walking speed.

During the measurement, it is important that participants do not speak, laugh or make any unnecessary movements since this will bias the estimated gait efficiency. This can be challenging, especially for young children and patients with intellectual disabilities. Moreover, this 6- to 12-minute walking test can already be exhausting for individuals that fatigue quickly due to their walking impairments. This is especially relevant when several different walking conditions need to be assessed (Kim et al., 2017). Conditions such as, walking with confection shoes, the conventional AFO and new aligned AFO. Sometimes patients are not even able to complete the full session.

### 1.4.2. Data processing

#### *1.1.1.1 Noisy breath-by-breath data*

At every exhale, a sample with the volumes of oxygen and carbon dioxide in the inhaled and exhaled air are measured during the gas analysis (Felt et al., 2015). From this *breath-by-breath* gas composition, the oxygen consumption and carbon dioxide production are computed. Data processing deserves caution since the data is sparsely sampled and noisy (Felt et al., 2015) and it is difficult to interpret the breath-by-breath data. The breath rate of the participant (typically 0.2-0.3 Hz during gait) determines the sample frequency and air volume in the expired breaths. This *inter-breath variability* introduces noise (Felt et al., 2015), see Figure 2. Children seem to have larger inter-breath variability in their respiratory response compared to adults (Armstrong & Barker, 2009). To reduce the effect of noise, the data is often averaged over multiple breaths. Typically, 1 up to 3 minutes are taken for this averaging of a *steady* period in the raw breath-by-breath, smoothed or filtered data (Felt et al., 2015; Kim et al., 2017).

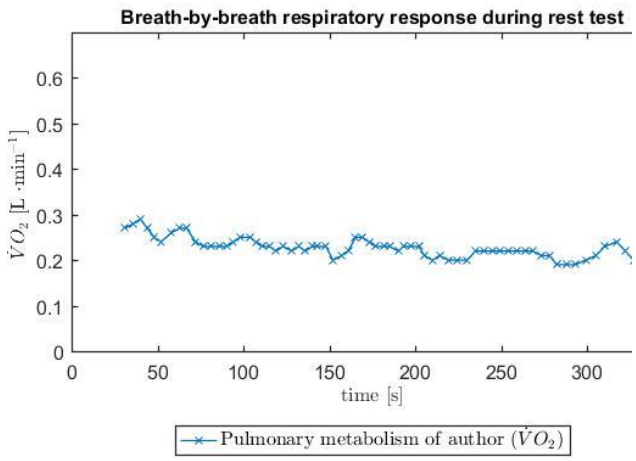



**Figure 2**. Illustration of noise in the breath-by-breath pulmonary metabolism ($\dot{V}O_2$) during a rest metabolism measurement of the author.

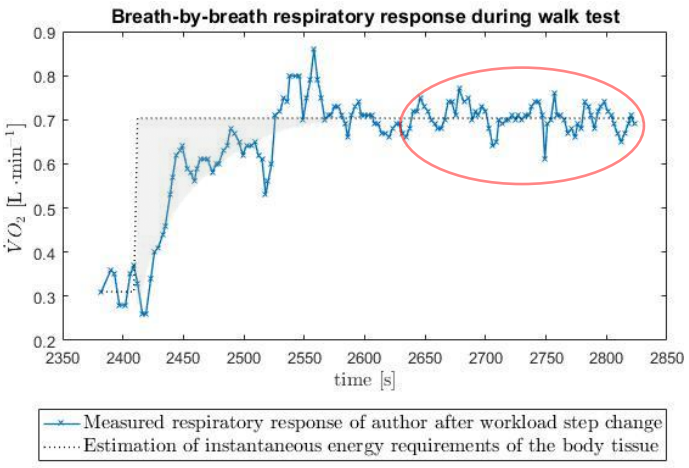



**Figure 3**. Illustration of the pulmonary respiratory response during a 6-minute walk test on a treadmill with a pre-defined constant speed. The dotted black line represents an estimation of the instantaneous metabolic energy demand of the body. The gray area illustrates the dynamic delay between the respiratory response and the instantaneous energy requirements (i.e. oxygen deficit). The red oval shows a steady period, or the so called steady-state metabolism.

#### *1.4.2.1. Steady-state metabolism*

There is a relatively large dynamic delay between the instantaneous energy consumption in the tissues and the respiratory response measured at the mouth, see Figure 3. Therefore, the mEE is often computed from an interval where a *steady-state* metabolism is reached. Steady-state refers to the state where upon the pulmonary metabolism reaches a constant level, where it equals the metabolism

of the tissue (Waters & Mulroy 1999). This state is reached after approximately 3 minutes of constant sub-maximal exercise (Armstrong & Barker, 2009).

This dynamic delay is also called *oxygen deficit* and partly originates from the slow mitochondrial dynamics (Selinger & Donelan, 2014; Kim et al., 2017). Moreover, the respiratory output does not instantaneously reflect the oxygen and carbon dioxide contents of the muscles due to a transit delay between the muscle and lungs and the lung ventilation frequency (Armstrong & Barker 2009). The lung ventilation frequency is neurally controlled and regulates the blood gases, making the respirator dynamics extra complex (Selinger & Donelan, 2014). Due to the neural control, the lung ventilation can rise before a movement takes place by anticipation on future exercise (Selinger & Donelan, 2014). Moreover, a breath sample only contains a part of the lung content and is not instantaneously reflecting the equilibrium in the blood gasses (Koller et al., 2017). "It is due to these complexities that energetic cost is traditionally only measured during long bouts of constant-intensity conditions" (Selinger & Donelan, 2014, *p*. 1406).

### *1.5. Wearable device optimization*

Energy expenditure of gait is recently used to optimize wearable assistive walking devices, to select the most efficient device parameters and assistance in alignment with the human operator. This upcoming technique is called "human-in-the-loop" optimization (Felt, Selinger, Donelan, & Remy, 2015; Koller, Gates, Ferris, & Remy, 2016; Kim et al., 2017; Zhang et al., 2017; Ding, Kim, Kuindersma, & Walsh, 2018; Slade et al., 2019). Within this technique the human is literally encompassed in the 'loop' and 'closes' this loop with one or multiple real-time physiological measures, such as mEE (Kim et al., 2017). This technique makes use of fast predictions of the mEE in combination with a smart search algorithm (e.g. 'gradient search' or 'Bayesian optimization') to explore different device parameters and their effect on the mEE during gait in a fast and efficient manner (Kim et al., 2017; Ding et al., 2018), see Figure 4.

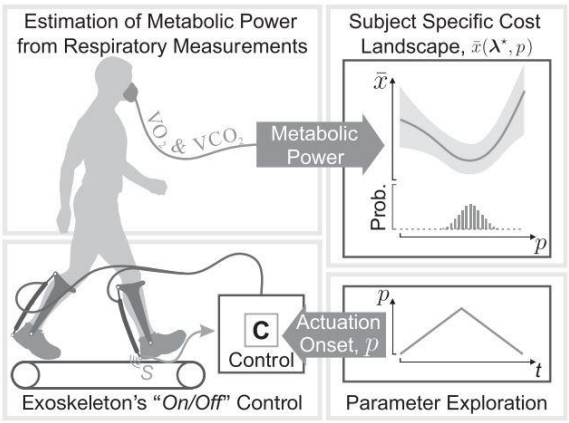



**Figure 4**. Illustration of human-in-the-loop optimization of a wearable ankle-foot exoskeleton. *Note:* Retrieved from Koller et al. (2017).

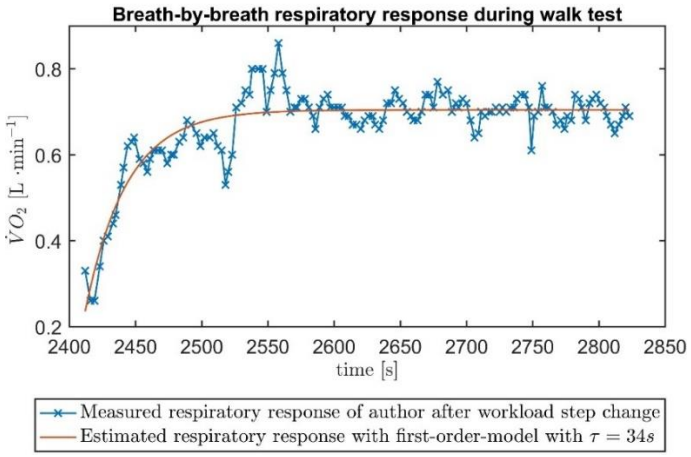



**Figure 5**. Illustration of the oxygen respiratory response of the author during a 6-minute walk test on a treadmill with a pre-defined constant speed together with the Instantaneous Cost Mapping (ICM) model estimation of the respiratory response with an identified subject-specific time constant of $\tau_{V\dot{O}_2} = 34s$.

### *1.6. Metabolic energy expenditure prediction*

The method to make fast predictions of the mEE in the human-in-the-loop techniques is also called 'Instantaneous Cost Mapping (ICM)'. The goal of the ICM technique is to predict the body's instantaneous metabolic demand, which is traditionally estimated by averaging the measured respiratory response during a steady-state metabolism (Koller et al., 2017). In this way, the mEE of a specific walking task can be predicted even before the steady-state metabolism is reached and the measurement duration can be reduced. The reported number of measured breaths that are required

to predict the instantaneous metabolic cost, vary between 40-60 breaths (Felt et al., 2015; Kim et al., 2017; Zhang et al., 2017; Ding et al., 2018). This corresponds to a measurement duration of approximately 2 to 3 minutes. This is less than half the conventional measurement duration of 6 minutes.

The underlying mechanism of the prediction is based on the theorem that the respiratory dynamics of the human body can be modeled as a simple first-order differential equation with an additional time constant ($\tau$), see Figure 5. The $\tau$ displays the time it takes to attain 63% of the change in the respiratory response to a step change in workload (Armstrong & Barker, 2009). This time constant can be identified on a subject-specific basis by fitting a curve to the respiratory response to a change in workload. However, Selinger and Donelan (2014) also report that the use of a general time constant can suffice to predict the instantaneous energy expenditure. They identified a general value of 42s for healthy adults which is already used among different human-in-the-loop studies (Kim et al., 2017; Zhang et al., 2017; Ding et al., 2018). A general time constant improves the practical applicability of the model since no subject-specific time constant need to be identified in advance.

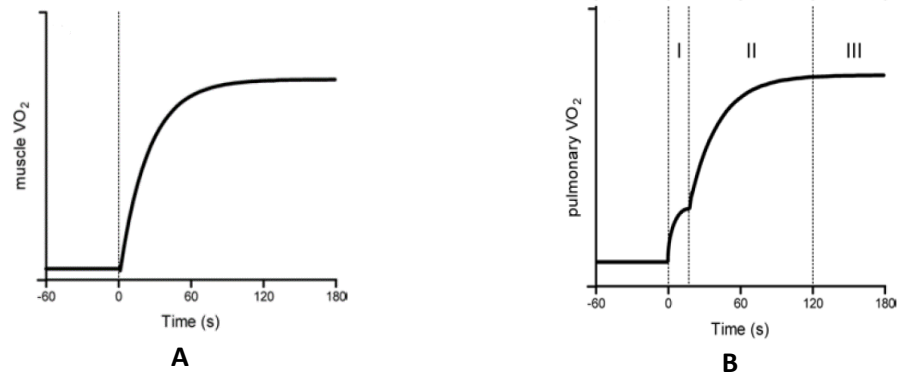



**Figure 6**. Schematic illustration of the A) the muscles and B) pulmonary rise in metabolic rate after a step change in work-rate. *Note:* retrieved from Armstrong, and Barker (2009) and calculations are based on the adopted results of Behnke et al. (2002) from experiments with rat skeletal muscles.

1.6.1. Evidence for the first-order time delayed model

There is evidence that, after a *step change* in workload, the oxygen consumption in the muscles ($m\dot{V}O_2$) is immediate and can be described as a mono-exponential time course during gait (Behnke et al., 2002; Armstrong & Barker, 2009), see Figure 6A. A step change in workload could be a change from rest to a moderate exercise intensity such as walking on a consistent speed. The pulmonary oxygen response ($\dot{V}O_2$) to such a change in workload, also referred as *oxygen kinetics,* can be described by a distorted curve with three phases (Potter & Unnithan, 2005; Armstrong & Barker, 2009;), see Figure 6B. The first phase reflects the *cardiodynamic* response, being the response of elevated cardiac output and increased blood flow (Potter & Unnithan, 2005; Armstrong & Barker, 2009). The second phase results from the arrival of the mixed venous blood at the lungs originating from the working muscles (Potter & Unnithan, 2005; Armstrong & Barker, 2009) and is followed by the third phase wherein a steady-state metabolism is reached and $\mathrm{m}\dot{V}O_2$ equals $\dot{V}O_2$.

The ICM model seems to neglect the distortions in the pulmonary oxygen response and focusses only on the second and third phase. This seems valid since there is a belief that the second phase of the pulmonary rate provides a faithful reflection of the underlying muscular respiration $\mathrm{m}\dot{V}O_2$ with a cohere of ± 10% (Whipp, Ward, Lamarra, Davis, & Wasserman, 1982; Potter & Unnithan, 2005; Barker et al., 2008; Armstrong & Barker, 2009). By fitting a curve to the second phase, the time constant of the muscle respiration can be estimated (Whipp et al., 1982; Potter & Unnithan, 2005). It should be kept in mind that this time constant is a simple model parameter and its fit does not represent the truly delay of the pulmonary respiratory response. Therefore, “its physiological relevance will be virtually meaningless due to the distortion of phase II of the response” (Armstrong & Barker, 2009, *p*.134).

Moreover, it should be noticed that the time constant found by a fit to the oxygen rate ($\tau_{\dot{V}O_2}$) results in a different time constant than that would be found from a fit to the mEE (in this study expressed as $\tau$). Potter and Unnithan (2005) stated that a $\tau_{\dot{V}O_2}$ of 30s is consistently reported in literature for the oxygen kinetics of adults. While, the human-in-the-loop techniques report a larger $\tau$ of 42s for healthy adult (Selinger & Donelan, 2014). It has been found that the oxygen kinetics in children is approximately 14% faster than that of adults with sub-maximal exercise (Fawkner, Armstrong, Potter, & Welsman, 2002; Armstrong & Barker, 2009).

### *1.7. Clinical relevance*

As far as known, the ICM model is not yet applied in adult and child patients with neurology and neuromuscular gait impairments, nor in typically developing children. It is unknown whether the model is useable for these populations. When the ICM model could replace the convention metabolic gas analysis, the time physicians need to assess the gait efficiency could be reduced and evaluations become less exhausting for patients that fatigue quickly. Moreover, application of the ICM model in rehabilitation could open ways to make human-in-the-loop protocols feasible for wearable walking device (i.e. AFO) tuning in patients.

### *1.8. Aim*

The primary aim of the present study is to investigate whether the ICM model, that make use of a first-order differential equation with an additional time constant $\tau$, could be used to reduce the measurement duration (< conventional 6 minutes) of metabolic gas analyses. The applicability of the ICM model will be tested on post hoc metabolic gas analysis data collected within the Amsterdam University Medical Centers (UMC) containing data of a variety of adult and child patients with gait impairments and healthy controls. First, the present study will evaluate whether a subject-specific ICM model correctly estimates the instantaneous mEE with the conventional measurement duration (6 minutes). Next, it aims to identify the group average estimation-error-measurement-duration relationship to explore on a group level whether the ICM model can reduce the measurement duration. Additionally, the present study will evaluate what effect this reduced measurement duration has on the correctness of the estimations on an individual level. Secondarily, the present study explores whether there are differences in the identified subject-specific time constants among the groups in the population and compare the identified $\tau$ values with the reported $\tau$ for healthy adults (42s) from literature. Because the use of a general time constant would improve the practical applicability of the ICM model, the effect of a general model on the measurement duration and its effect on the individual correctness will additionally be explored.

### *1.9. Hypothesis*

It is expected that the cardiorespiratory system of the patients with neurology and neuromuscular diseases are not affected. Therefore, it is hypothesized that the ICM model can correctly describe their respiratory response and reduce the conventionally measurement duration (6 minutes) to a similar duration as used within the human-in-the-loop techniques of 2-3 minutes. It is hypothesized that the use of a general ICM model would increase the estimation errors of the model compared to a subject-specific model. Resulting in a lower individual correctness rate and a longer required measurement duration. Nevertheless, it is expected that the general model is still capable of reducing the conventional measurement duration.

# 2. Methods

## *2.1. Data Collection*

### 2.1.1. Data providers

Data collection took place from April until July 2019 within two rehabilitation departments of the Amsterdam UMC, location Amsterdam Medical Center (AMC) and Free University medical center (VUmc) located in Amsterdam. The data originates from 7 previous performed research studies and regular clinical service measured in 2006 until 2019 with varying protocols and pathologies.

### 2.1.2. Approval

Since the present study did not require extra experimental measurements, permission for this study was given internally within the rehabilitation departments of the hospitals. All participants approved data collection and signed an informed consent within the previous performed studies or gave approval for using the data collected during clinical service available for research. Moreover, a data sharing agreement was signed by the legal department of the AMC hospital to get permission to work with the data of both hospitals within the protected work environment at the VUmc location.

### 2.1.3. Study population

Two of the 7 studies (i.e. MOVING-MS, MSWALK-I) include adult MS patients ($n$ = 19) and matched healthy adult (HA) subjects ($n$ = 19). One study includes young healthy adults only ($n$ = 12). Three studies contained typically developing (TD) children ($n$ = 30). One study (i.e. PROOF-AFO) includes adult patients ($n$ = 37) with a variety of neuromuscular disorders (ND; i.e. $n$ = 8 poliomyelitis, $n$ = 16 Charcot-Marie-Tooth disease, $n$ = 1 inclusion body myositis, $n$ = 2 myotonic dystrophy, $n$ = 8 peripheral nerve injury, $n$ = 1 chronic inflammatory demyelinating polyneuropathy, $n$ = 1 miyoshi distal myopathy, $n$ = 1 partial paraplegia) “with non-spastic paresis or weakness of the calf muscles” (Waterval et al., 2017, $p$. 2). Additionally, data of children with CP ($n$ = 31), measured during clinical service was collected.

### 2.1.4. Protocols of the performed studies

#### *2.1.4.1. Treadmill studies*

Both MS studies were performed on the GRAIL© treadmill system of MotekFore Link B.V. located at the VUmc. The MS patients of the MSWALK-I study performed a 12-minute walk test (MWT) with a self-paced comfortable walking speed. Two follow-up measurements were performed after 2-3 and 5-6 months. The healthy matched control subjects only performed the 12-MWT once.

The MS patients of the MOVING-MS study performed twice a 6-MWT with two different pre-defined walking speeds within one trial. The healthy control subjects of this study performed 3 times the 6-MWT at 3 different pre-defined walking speeds.

#### *2.1.4.2. Overground studies*

In all other studies and clinical service (Bolster, Balemans, Brehm, Buizer, and Dallmeijer, 2017; Waterval et al., 2017) the participants performed a 6-MWT on their comfortable self-paced walking speed over a marked oval over-ground track, of either 34, 35 or 40 meters, which was marked on the ground by small dots with indications of the distance.

#### *2.1.4.3. Walking condition*

In most protocols, participants walked on confection shoes without additional walking aids. However, the CP children that were measured during clinical service might have used their regular walking aids and/or orthotics when needed. Moreover, in the protocol of the PROOF-AFO study, 8 conditions were tested within one subject at 3 different points in time: (T1) confection shoes, previous AFO, (T2) 5 conditions with differing stiffnesses of the AFO, and a 3-month follow up measure for (T3) the in T2

identified most economical stiffness setting of the AFO. This study also allowed subjects to walk with additional helping aids such as a stick, crutch, or wheel walker when needed.

*2.1.4.4. Used indirect calorimeter*

The metabolic gas analysis in the previous studies were performed with different indirect calorimeters. The used devices at the VUmc are the METAMAX® 3B Cortex Biophysik (Leipzig, Germany) and the COSMED® K5 (Rome, Italy). The device used in the AMC was one previous version of this later device, the COSMED® K4B$^2$, (Rome, Italy).

## *2.2. Research outcome measures*

### 2.2.1. Primary research outcome

The primary research objective of the present study was to explore whether the ICM model can reduce the conventional measurement duration (6 minutes) of the metabolic gas analysis in adult and child patients with a variety of neurology and neuromuscular diseases with subsequent gait impairments and healthy controls. To investigate this, the present study evaluated whether a subject-specific model could correctly estimate the instantaneous energy expenditure on an individual and average basis. Next, it investigated whether the measurement duration could be shortened by evaluating the average estimation-error-measurement-duration relationship of the subject-specific model. Finally, it evaluated what effect this reduced measurement duration has on the individual correctness of the estimations.

### 2.2.1. Secondary research outcomes

Secondarily, the present study evaluated the goodness-of-fit of the ICM model curve. It investigated whether the identified subject-specific $\tau$ differs among the groups in the included population. It tested whether the average identified $\tau$ differs from the general $\tau$ reported for healthy adults (42s; Selinger & Donelan, 2014). Finally, the present study evaluated what effect the use of a general $\tau$ has on the individual correctness of the estimations, the required measurement duration and how this shortened measurement duration effects the individual correctness of the estimations.

## *2.3. Inclusion criteria*

Since this study is the first exploratory study that evaluates the clinical application of the ICM model in both adult and child patients, strict inclusion criteria were handled to include only 'clean' data that approximates a step change in workload since the ICM model currently only fit the respiratory response to a step change in workload. Therefore, the data was only included when:

1. an unfiltered breath-by-breath version was available.
2. a steady-state metabolism was reached according the statistical steady-state determination method of Schwartz (2007), see paragraph 2.4.1 below.
3. the participant started from a steady rest metabolism, see paragraph 2.4.2 below.
4. the walking speed was consistent and did not differ more than 10% of the average walking speed over the duration of the walking trial, nor showed a decreasing or increasing trend in the walking speed. See paragraph 2.4.3. below.

When all above mentioned inclusion criteria were met within multiple trials of the subject, only one trial was included per subject to prevent distortion of the generalizability.

## *2.4. Inclusion procedure*

The mathematical calculations were performed with the software program MATLAB® (R2017b; MathWorks, Natick, Massachusetts, USA).

2.4.1. Steady-state determination

The present study used the statistical approach of Schwartz (2007), applying a non-parametric correlation, Kendall's tau ($\tau_b$) to determine whether the participants reached a steady-state metabolism for inclusion and to determine the 'true' steady-state mEE as control value. This approach used a moving time window. The statistic $\tau_b$ was used to decide whether there is a rising, falling or random trend within the window and can be considered as steady or unsteady (Schwartz, 2007). This make use of a computation of the subsequent rises ($Q$) counted from each point forward through the window until the second last point (Schwartz, 2007), see equation (1).

$$Q = \sum_{i=1}^{n_w-1} \sum_{j=i+1}^{n_w} \begin{Bmatrix} 1 & \text{value}_j > \text{value}_i \\ 0 & \text{otherwise} \end{Bmatrix} \quad (1)$$

With $n_w$ representing the number of datapoints within the time window and $i$ and $j$ being indexes of the respiratory data $y$ to assess whether $y_j > y_i$ . Then, $\tau_b$ can be defined as,

$$\tau_b = \frac{4Q}{n_w(n_w-1)} - 1 \quad (2)$$

with the standard deviation of $\tau_b$ being,

$$\sigma_{\tau_b} = \frac{2(2n_w+5)}{9n_w(n_w-1)} \quad (3)$$

For a steady time window, the expected value $\mathrm{E}(\tau_b)$ is 0 and the $z$-score for $\tau_b$ can be computed as,

$$z_{\tau_b} = \frac{\left(\tau_b - \mathrm{E}(\tau_b)\right)}{\sigma_{\tau_b}} = \frac{(\tau_b - 0)}{\sigma_{\tau_b}} \quad (4)$$

from which the probability $p$-value can be computed to test the null hypothesis ($H_0$: data within the time interval is steady). The rejection level of $H_0$ to assess the steadiness of the window was set to a level of $\alpha = .10$, similar as Schwartz (2007). This rejection level determines whether the interval is classified as steady ($p > .1$) or as rising/falling ($p < .1$).

The length of the time window for the walking test was set to a length of 180s similar as Schwartz (2007), a smaller window of 90s was used for the rest period since the rest period length differed per trial (being less than 4.5 minutes in some cases), becoming too short evaluation of a moving window of 180s. Especially taken into account that half the size of the window length at the beginning and end of the dataset were not assessed for steadiness, see Figure 7. This consideration seems valid since the rest measurement is expected to be more stable compared to the walking test. When no steady-state windows were detected in either the rest or walking test, the trial was excluded from analysis.

Different than reported in the study by Schwartz (2007), the present study calculated the average of the window resulting in the moving average ($M_{mov}$) of every time window. Caution should be taken by calculating the mean of the window since metabolic gas analysis data is sampled with a varying sample frequency. Therefore, a regular $mean$ function, implemented in the MATLAB software, could not be used. The integral of the time window should be used instead, see equation (5).

$$M_{mov} = \frac{\int_{t_1}^{t_2} \dot{V}O_2(t)dt}{t_2 - t_1} \quad (5)$$

With $t_1$ and $t_2$ representing the borders of the time window. When a window was classified as stable, the moving average of that specific window was taken together with the $M_{mov}$ of the other steady windows to determine the 'true' measured steady-state mEE control value. In both the rest ($mEE_{rest}$) measurement as in the walking trial ($mEE_{walk}$), see equation (6). With $n_{sw}$ representing the number of steady windows.

$$mEE = \frac{\sum_{i=1}^{n_{sw}} M_{mov,i}}{n_{sw}} \quad (6)$$

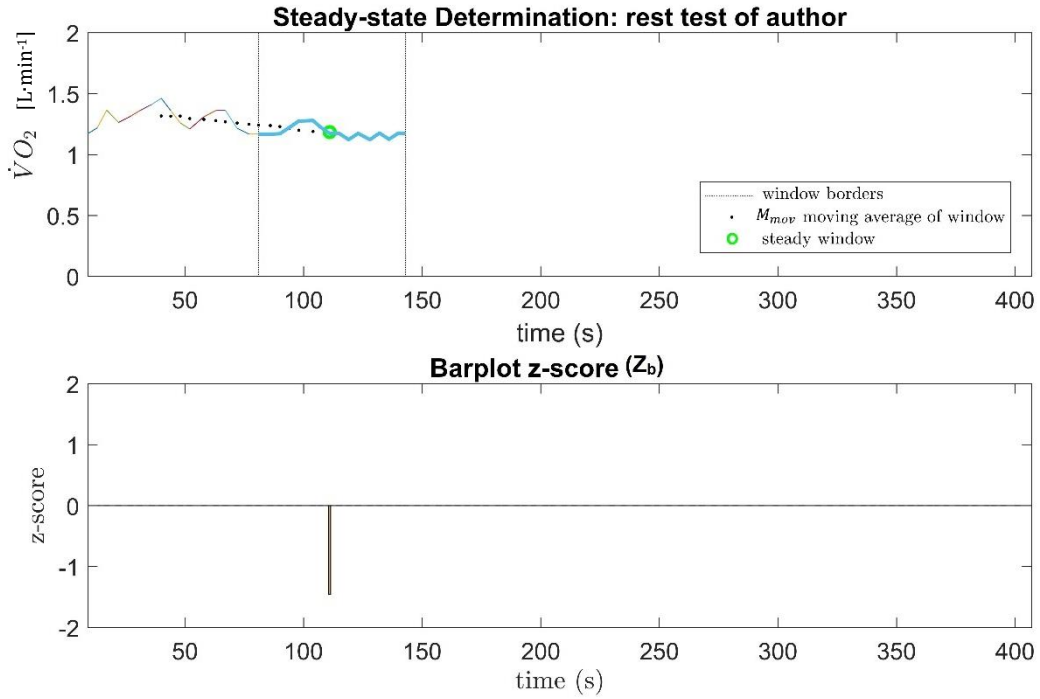



**Figure 7**. Illustration of the steady-state determination approach based on the statistical method of Schwartz (2007). The upper graph shows the oxygen rate $\dot{V}O_2$ [L·min$^{-1}$] of a rest measurement performed by the author. The thin vertical lines in the upper graph represent the window borders of the current window evaluated (thick blue line) for steadiness. The black dots represent the moving average ($M_{mov}$) values of the evaluated windows and turn into green circles when a window is classified as steady. The lower graph represents the $z$-scores ($z_{\tau_b}$) of the observed Kendall's tau ($\tau_b$) when a window is considered to be steady. It can be seen that the first data points within half the window length (in this case 45s) are not assessed for steadiness.

2.4.2. Rest metabolism criteria

To evaluate whether the participants started the walk test from a rest metabolism, the first 10 datapoints ($y_{1:10}$) of the walking test were assessed to meet equation (7).

$$mEE_{rest} - 2 \cdot s_{rest} \leq y_{1:10} \leq mEE_{rest} + 2 \cdot s_{rest} \quad (7)$$

Where $s_{rest}$ represents the standard deviation of the rest metabolism and is calculated as follows:

$$s_{rest} = \sqrt{\frac{\sum_{i=1}^{n_r}(y_i - mEE_{rest})^2}{n_r - 1}} \quad (8)$$

Where the $n_r$ represent the number of datapoint in the total rest measurement. See Figure 8 for an illustrative overview of the steady-state and rest metabolism inclusion criteria.

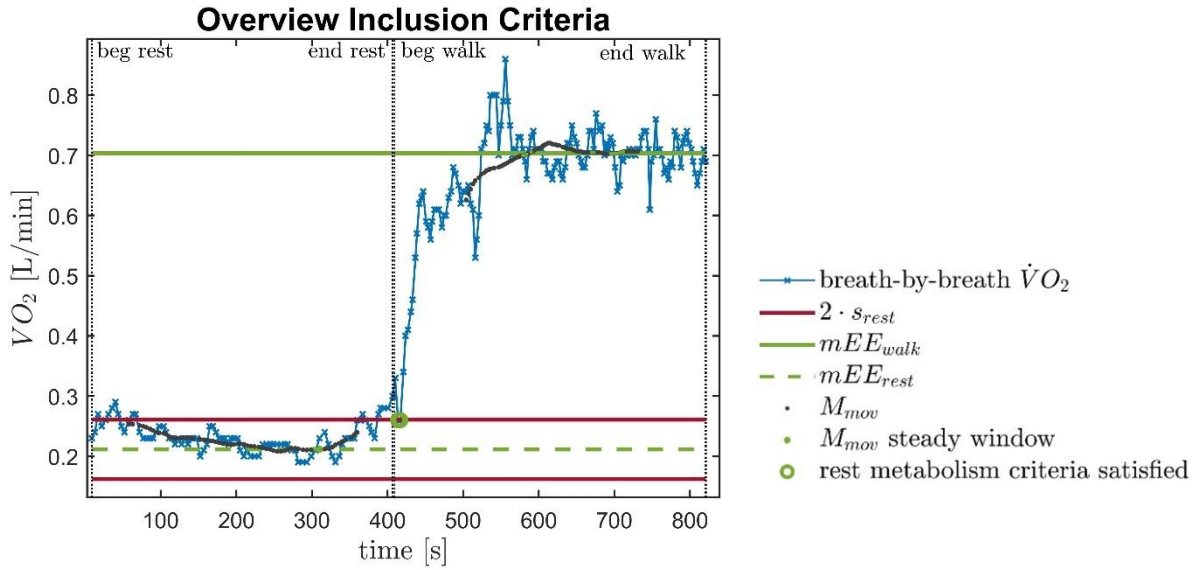



**Figure 8**. Illustration of the inclusion criteria 2 and 3 concerning the steady state metabolism in the rest and walk measurements and starting from a rest metabolism in one of the first 10 data points of the walk measurement. Data is collected during a pilot measurement with the author.

2.4.3. Walking speed criteria

In the trials that took place over-ground, the time upon a lab was completed (i.e. 'lap-times') was recorded and written down in a timetable (not digitalized). These lap-times were used to compute the average speed over one lap to assess the consistency in walking speed among the completed laps. In a part of the walking trials the timetable was not available, and the consistency of the walking speed could not be evaluated and were therefore excluded from analysis. The lab-speed ($v_{lap}$) was computed from the duration of the lab ($\Delta t_{lab}$) and walked distance ($\Delta d_{lab}$) within the lab, varying from 34 to 40 meters depending on the used track across studies, see equation (9).

$$v_{lap} = \frac{\Delta d_{lap}}{\Delta t_{lap}} [m/s] \quad (9)$$

For the MSWALK-I study performed on the GRAIL, with the self-paced preferred walking speed, the consistency of walking speed was computed from the previous calculated average walking speed per minute that was provided with the data. These values were calculated from the measured treadmill belt speed of the GRAIL, see equation (10).

$$v_{minute} = \sum_{i=1}^{n_s} \frac{v_{GRAIL\ signal,i}}{n_s} [m/s] \quad (10)$$

With $n_s$ representing the number of individual samples of the measured GRAIL signal. From the $v_{minute}$ or $v_{lab}$ the average walking speed $v_{walk}$ was calculated.

To assess the consistency of the walking speed, a first-order linear regression was performed to detect an increasing, falling or steady trend within the $v_{minute}$ or $v_{lab}$ values. The absolute slope ($\beta_1$) of the regression model [m·s$^{-1}$·$n_v$$^{-1}$] had to be less than 10% of the $v_{walk}$ over the total measured laps or minutes ($n_v$), see equation (11).

$$|\beta_1| < \frac{10\% \cdot v_{walk}}{n_v} \quad (11)$$

Additionally, the absolute deviance ($d_v$) of the $v_{minute}$ or $v_{lab}$ from $v_{walk}$ was computed for each minute or lap and had to be also less than 10% of $v_{walk}$, see equation (12).

$$|d_v| < 0.10 \cdot v_{walk} \quad (12)$$

Both methods were needed since positive and negative $d_v$ could outweigh each other, potentially resulting in a regression that satisfy equation (11) while equation (12) was not satisfied. Visa versa, when $d_v$ satisfied equation (12) a rising or falling trend could still exceed the criteria of equation (11).

## *2.5. Main analysis*

### 2.5.1. Instantaneous cost mapping

As far as known, Whipp et al. (1982) were the first researchers that aimed to model non-steady-state metabolic gas dynamics, in their case, during cycling. They found that for some conditions, the mEE could be described by a first-order differential equation with an additional time delay (Whipp et al., 1982; Selinger & Donelan, 2014). Later on, Selinger and Donelan (2014) used this model as starting point to model the relationship between the instantaneous mEE ($\dot{E}$) and the measured respiratory response ($y$) measured at the mouth, see Figure 9.

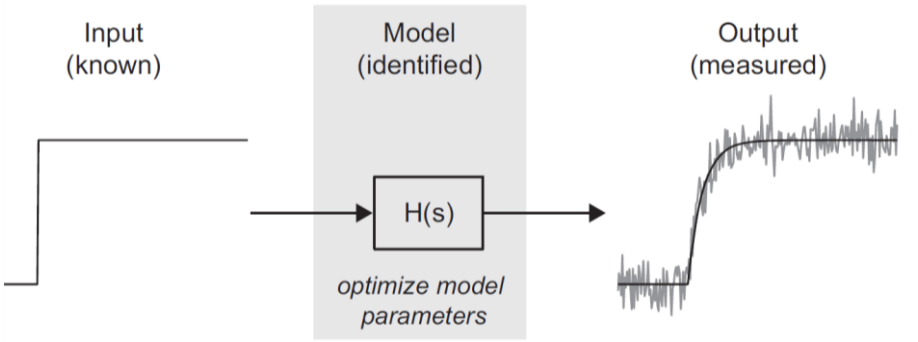



**Figure 9**. Schematic view of the modeled relationship between instantaneous metabolic energy expenditure (input) and measured respiratory response (output). *Note:* retrieved from Selinger and Donelan (2014).

This identified model can be expressed in the Laplace domain as:

$$Y(s) = H(s)X(s) \quad (13)$$

Where the transfer function of $H(s)$ has the form,

$$H(s) = \frac{\kappa}{\tau s + 1} e^{-\delta s} \quad (14)$$

$X(s)$ represents input instantaneous energy expenditure $\dot{E}$, and $Y(s)$ the output measured response $y$. $H(s)$ represents the identified model with the additional parameters $\tau, \delta$ and $\kappa$. The model parameter $\tau$ is the time constant describing the respiratory dynamics and can differ among subjects. The model parameter $\delta$ represents a fixed time delay between the muscle and pulmonary metabolism (Selinger & Donelan, 2014). The parameter $\kappa$ reflects the ratio between the amplitude change in the instantaneous energy expenditure and the measured respiratory response as consequence of the change in external workload. In the case of a steady-state metabolism the input $X(s)$ equals the output $Y(s)$ and therefore $\kappa$ = 1. Selinger and Donelan did not found values for $\delta$ that were discernible form zero (within a sample size of $n$ = 10) and found that the simplified version of equation (14) could well approximate $\dot{E}$ during gait, see equation (15).

$$H(s) = \frac{1}{\tau s + 1} \tag{15}$$

Later on, other research groups (Felt et al., 2015; Koller et al., 2016; Malcolm et al., 2017; Zhang et al., 2017; Ding et al., 2018) have used this identified model to make a fast predictions of instantaneous energy expenditure from the non-steady-state metabolism for the human-in-the-loop wearable device optimization, see Figure 10. This technique replaces the traditional steady-state energy expenditure determination.

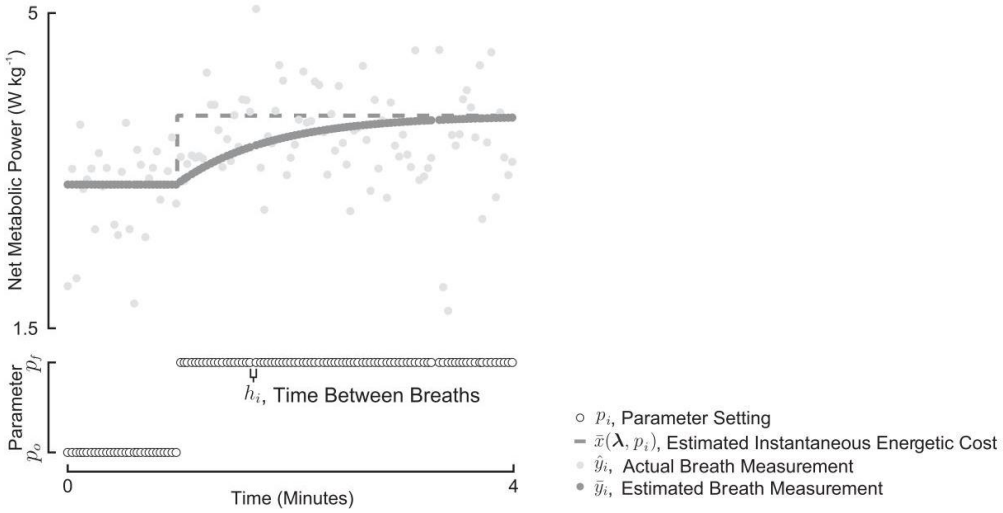



**Figure 10**. A representative example of the actual measured breath-by-breath output (light gray dots) together with the estimated metabolic response (dark gray dots) and the estimation of the instantaneous energy expenditure (dark gray dots) given a step change in actuation onset timing for a set of pneumatically powered ankle exoskeletons. The estimation is made with the first-order model with a subject-specific identified time delay of $\tau = 51.9$ s. *Note:* retrieved from Koller et al. (2017)

The discretized step response of the model on a breath-by-breath basis can be described as the following (Felt et al., 2015; Koller et al. 2016; Zhang et al., 2017):

$$\overline{y}_{i+1} = \left(1 - \frac{dt_i}{\tau}\right)\overline{y_i} + \frac{dt_i}{\tau}\,\bar{\dot{E}}. \tag{16}$$

With $dt_i = t_{i+1} - t_i$, the time difference between breaths sample $i$ and the next sample. Where, $\bar{\dot{E}}$ is the unknown estimated instantaneous energy expenditure and $\bar{y}_{i+1}$ the estimated respiratory response of an individual breath that depend on the previous breath measurement $i$.

A solution for $\bar{\dot{E}}$ can be found by minimizing the sum of squared error between the estimated $\bar{y}_i$ and the actual measurements $\hat{y}_i$ (Koller et al., 2017). This can mathematically be described as:

$$\min \sum_{i=1} (\bar{y}_i - \hat{y}_i)^2. \tag{17}$$

This can be solved by expressing equation (16) in a system of linear equations (Felt et al., 2015; Koller et al., 2015):

$$\begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{bmatrix} = \mathbf{A} \begin{bmatrix} \bar{y}_1 \\ \bar{\dot{E}} \end{bmatrix} \tag{18}$$

in which $\bar{y}_1$ is the second unknown variable, representing the initial respiratory response value. Matrix $\mathbf{A}$ can be constructed by (Felt et al., 2015; Koller et al., 2016; Koller et al., 2017; Zhang et al., 2017):

$$\mathbf{A} = \begin{cases} 1 & i = 1, j = 1 \\ 0 & i = 1, j = 2 \\ \mathbf{A}_{i-1,j} \cdot \left(1 - \frac{dt_i}{\tau}\right) & i > 1, j = 1 \\ \mathbf{A}_{i-1,j} \cdot \left(1 - \frac{dt_i}{\tau}\right) + \frac{dt_i}{\tau} & i > 1, j = 2 \end{cases} \quad (19)$$

The rows of $\mathbf{A}$ correspond to the individual breath ($i$) samples (Koller et al., 2017). The first column ($j = 1$) corresponds to the solution of $\bar{y}_1$, and the second column ($j = 2$) to the solution of $\bar{\dot{E}}$. This matrix notation allows an efficient method to find the optimal solution for $\bar{y}_1$ and $\bar{\dot{E}}$ that satisfy equation (17) (Felt et al., 2015) by, using, $\boldsymbol{A}^+ = (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} \boldsymbol{A}^\intercal$, the pseudoinverse of $\boldsymbol{A}$, as follows:

$$\begin{bmatrix} \bar{y}_1 \\ \bar{\dot{E}} \end{bmatrix} = \boldsymbol{A}^+ \bar{y} \quad (20)$$

In this way, $\bar{y}_1$ and $\bar{\dot{E}}$ will be optimized for the least squared-error possible (Felt et al., 2015; Zhang et al., 2017). Felt et al. (2015) and Zhang et al. (2017) have performed this method in a custom-made MATLAB function and made this available online which is used for the analysis of the present study.

#### 2.5.2. Unit of the mEE

In literature, many different terms are used to express mEE. From metabolic work or power (Waters & Mulroy, 1999) to oxygen- or energy- rate, cost, and consumption (Baker et al., 2001; Brehm et al., 2007). All terms slightly differ in computation. The choice of calculation has influence on both the model parameters and the model outcome. This should be considered when comparing the results with others. The present study compared different methods, but for clarity only the result of the mEE calculated according the equation of Garby and Astrub (1987) will be reported, see equation (21). See Appendix 9.1 for a detailed description of different calculation methodologies.

$$mEE = (4.960 \cdot RER + 16.040) \cdot \dot{V}O_2 \quad (21)$$

#### 2.5.3. Subject-specific $\tau$ model parameter optimization

To solve for the subject-specific model parameter ($\tau$), the discretized first-order time response of equation (16) has been used to fit the respiratory response expressed in energy consumption [J·kg$^{-1}$·min$^{-1}$]. The identified steady-state values of $mEE_{rest}$ and $mEE_{walk}$ are used for the initial respiratory response $y_1$ and instantaneous energy expenditure $\dot{E}$, respectively. In this way $\tau$ was the only unknown model parameter to be identified. Then, a least-squares gradient search was used to minimize the residuals between the model $\bar{y}$ and the measured data $\hat{y}$ until the first steady window was reached. The optimization was performed with the implemented MATLAB function $lsqnonlin$ of the Optimization Toolbox. The lower search boundary was defined as the duration of the time interval between the first two breaths of the walking trial. The upper search boundary was set to a fixed value of 240s, assuming that a steady-state metabolism should be reached within 4 minutes of consistent exercise. Since only one model parameter needed to be optimized a 1-Dimensional *line search* could also have been used for this optimization problem. However, due to computational reasons the gradient search was preferred. It is unlikely for this specific optimization problem that the solution of the gradient search will result in a local minimum. However, to test this, different initial starting points within the search boundaries were explored using a test trial and all solutions end up in the same result.

## *2.6. Evaluation of the correctness of the model outcome*

### 2.6.1. Individual correctness of $\bar{\dot{E}}$

The ICM model outcome $\bar{\dot{E}}$ was considered to be correct on an individual basis when it lies within the 95% confidence interval ($95\%CI_{meas}$) boundaries of the statistically determined steady-state $mEE_{walk}$. The $95\%CI_{meas}$ was computed from the 95% confidence intervals of separate steady time windows. Since the number of samples within a time windows is small the $95\%CI_{meas}$ was calculated using the critical value of the *t*-distribution ($T_{df}$) according to Field (2013) of a two-tailed test with a probability level of $p$ = .05, for the appropriate *degrees of freedom* ($df_i = n_{w,k} - 1$), with $n_{w,k}$ representing the number of datapoints within the $k^{\text{th}}$ steady window, see equation (22).

$$95\%CI_{meas} = \frac{\sum_{k=1}^{n_k} M_{mov,k} \pm (T_{df,k} \cdot \sigma_k)}{k} \quad (22)$$

Where $n_k$ represents the number of steady windows and $\sigma_k$ represents the standard error of the mean of steady window $k$, and can be calculated as,

$$\sigma_k = \frac{s_k}{\sqrt{n_{w,k}}} \quad (23)$$

With $s_k$ the standard deviation of the individual window $k$, see equation (24).

$$s_k = \sqrt{\frac{\sum_{i=1}^{n_{w,k}} (y_{i,k} - M_{mov,k})^2}{df_k}} \quad (24)$$

Where $y_{i,k}$ represents the individual datapoints within the $k^{\text{th}}$ time window.

### 2.6.2. Equivalence of average $\bar{\dot{E}}$ and $mEE_{walk}$

To evaluate whether the ICM model outcome is equivalent to the statistically determined steady-state control value on an average group level, a one-sample Equivalence Test is performed. The clinical equivalence value was set at $\Delta_E = 13\% \cdot M_{mEE_{walk}}$ based on the *clinical relevant difference* for mEE (Buckon et al., 2004), with $M_{mEE_{walk}}$ the group average steady-state control value. The 95% confidence interval was computed with the critical value of the *t*-distribution and the standard error of the mean of the ICM model outcome $\bar{\dot{E}}$.

### 2.6.3. Goodness-of-fit model curve $\bar{y}$

Similar as Selinger and Donelan (2014) the goodness-of-fit between the model $\bar{y}$ and the measured data $\hat{y}$ was assessed by calculating the $R^2$ value. The $R^2$ can be calculated using the model sum of squares ($SS_M$) divided by the total sum of squares ($SS_T$) according to Field (2013), see equation (25). The $R^2$ value represents the percentage of variation in $\bar{y}$ that can be explained by the model (Field, 2013).

$$R^2 = \frac{\sum_{i=1}^{n_m} (mEE_{walk} - \bar{y}_i)^2}{\sum_{i=1}^{n_d} (mEE_{walk} - \hat{y}_i)^2} = \frac{SS_M}{SS_T} \quad (25)$$

With $n_m$ representing the number of datapoints in the estimated model $\bar{y}$, and $n_d$ representing the number of datapoints in the measured data $\hat{y}$.

### *2.7. Evaluation of the required measurement duration*

As mentioned in the introduction, the human-in-the-loop techniques measure 40-60 breaths (Felt et al., 2015; Kim et al., 2017; Zhang et al., 2017; Ding et al., 2018), that correspond to approximately 2 to 3 minutes to estimate the mEE. Zhang et al. have described their considerations. They chose to select 2 minutes based on an acceptance level of 4% for the mean estimation-error-measurement-duration relationship, see Figure 11. They mentioned that this level "seemed to strike a good balance between estimation errors (an average of 4% error) and trial duration" (*p*. 9, Zhang et al., 2017 [supplementary material]).

The present study adopted this acceptance level of 4% estimation error as starting point to explore whether the ICM model can reduce the measurement duration with respect to the conventional 6 minutes in the included population. However, as can be seen from Figure 11, still many trials have estimation errors above 4% at a measure duration of 2 minutes, resulting in less correct estimations on an individual level. Therefore, the present study identified the required measurement duration ($\Delta t_{meas}$) based upon a somewhat stricter level, by taking the point where upon the upper boundary of a 95% confidence interval (UBCI) of the group estimation-error-measurement relationship reaches the 4% of estimation error.

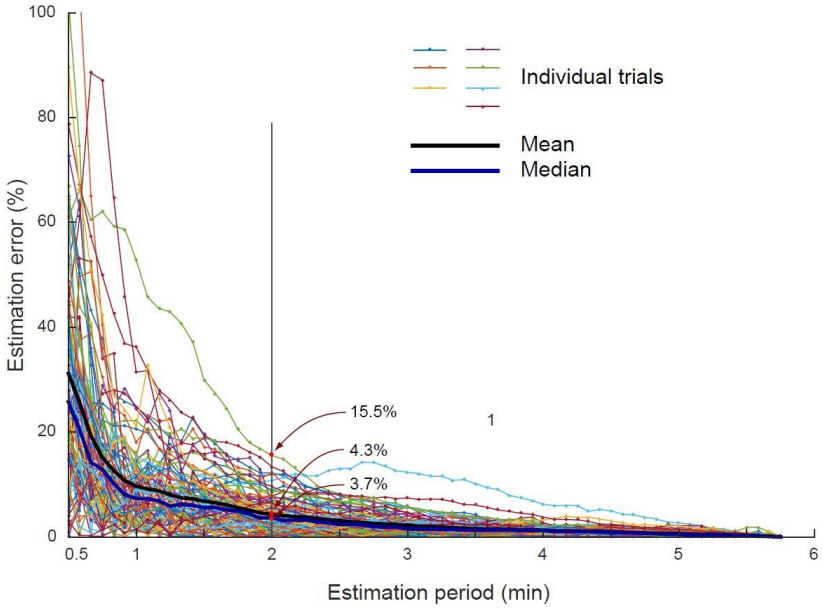



**Figure 11**. Estimation-error-measurement-duration relationship of the validation trials in healthy adults using a general $\tau$ of 42s of the study of Zhang et al. (2017). They calculated the metabolic energy expenditure according to the equation of Brookway (1987). *Note*: retrieved from Zhang et al. (2017) [supplementary materials].

#### 2.7.1. Estimation-error-measurement-duration relationship

The individual estimation-error-measurement-duration relationship was computed by calculating the percentage of absolute error between the model outcome and the steady-state control value, see equation (26). Computed for all available breath samples in the data (full post hoc measurement duration) up until solely the first sample. The group average of the estimation-error-measurement-duration relationship was computed by taken the mean of all interpolated individual estimation-error-measurement-duration relationships to a time vector of 00m:30s to 06m:00s. The group 95% confidence interval of the relationship was again computed with the critical value of the t-distribution and standard error of the group estimation-error-measurement-duration relationship.

$$\%\ estimation\ error = \frac{|\bar{\dot{E}} - mEE_{walk}|}{mEE_{walk}} \cdot 100\% \qquad (26)$$

### *2.8. Statistical analysis*

Statistical analysis was performed in SPSS statistical software system (SPSS Inc., Chicago, II, USA; version 24.0). Descriptive statistics were used to summarize demographic variables and to calculate the group mean ($M$), standard deviation ($SD$), standard error of the mean ($SE$), minimal and maximal values and/or median ($Mdn$) of $\dot{\bar{E}}$, $\tau$, and $R^2$values.

The distributions of the $\dot{\bar{E}}$, $\tau$, and $R^2$ values were evaluated with the skewness, kurtosis, Kolmogorov-Smirnov (K-S) and Levence test values. Based on this, a distinction between parametric and non-parametric tests was made. The significance level was set at $p$ = .05, two-tailed.

The effect size of the one-sampled T-test is expressed by Cohen's $d$ and computed by dividing the $t$-statistic value by the square root of number of observations (Field, 2013). The effect size of the Independent-Samples Man-Whitney U test was computed from the test $z$-score divided by the square root of the total number of observations (Field, 2013).

# 3. Results

### *3.1. Collected data*

A total of 169 subjects containing 482 walking trials originating from 8 different data sources have been tested for the inclusion criteria. Based on the first criteria, stating that a raw breath-by-breath version of the data must be available, a total of 18 subjects including 205 walking trials were excluded ($n$ = 151, trials = 277). Additionally, due to a lack of availability of walking speed information a total of 48 subjects and 77 trials were excluded ($n$ = 103, trials = 200). Based on the second and third inclusion criteria, stating that the walking trial must start from a rest and reach a steady-state metabolism, a total of 52 subjects and 118 trials were excluded ($n$ = 51, trials = 82). Based on the last criteria, stating that the walking speed needs to be consistent over the walking trial, a total of 23 subjects and 34 trials were excluded. Resulting in a total of 28 subjects containing 48 walking trials that were suitable for analysis among which 5 children and 23 adults. Only one trial per subject was included for analysis.

### *3.2. Descriptive data*

| **Table 1. Descriptive data** | | ***n* = 28** | ***n* = 23** | ***n* = 5** |
|---|---|---|---|---|
| | | *Total* | *Adult* | *Children* |
| Gender male/female | [n] | 13/15 | 10/13 | 3/2 |
| Age | [years] | 45.3 ± 21.9 (7-75) | 52.9 ± 15.7 (18-75) | 10.2 ± 2.2 (7-12) |
| Body Mass | [kg] | 71.7 ± 21.8 (25-103) | 79.8 ± 14.1 (53-103) | 34.7 ± 6.0 (25-41) |
| Body length | [cm] | 168.1 ± 16.5 (127-192) | 174.1 ± 10.0 (154-192) | 140.5 ± 11.3 (127-157) |
| BMI | [kg/m²] | 24.8 ± 5.8 (14.2 – 43.2) | 26.4 ± 5.0 (17.9-43.2) | 17.6 ± 2.8 (14.2 – 21.0) |
| Pathology | [n] | | | |
| *Healthy control* | | *5* | | |
| HA | | | 1 | |
| TD | | | | 4 |
| *Patients* | | *23* | | |
| MS | | | 8 | |
| ND | | | 14 | |
| CP | | | | 1 |

All data are presented by the mean ± standard deviation (range) unless otherwise stated. Abbreviations: $n$ = number of participants, HA = Healthy Adults; TD = Typical Developing children; MS = Multiple Sclerosis; ND = a variety of neuromuscular diseases; CP = Cerebral Palsy.

### *3.3. Primary outcome measure*

#### 3.3.1. Correctness of the ICM model outcome $\bar{\dot{E}}$ for conventional measure duration

##### *3.3.1.1. Individual correctness of $\bar{\dot{E}}$*

For the total conventional measure duration of 6 minutes the ICM model made in 27 out of 28 (96%) trials a correct estimation on an individual level. This implies that the instantaneous energy expenditure $\bar{\dot{E}}$ lie within the individually determined $95\%CI_{meas}$ control boundaries and could be interpreted as correct estimations of the instantaneous mEE representing the steady-state energy expenditure $mEE_{walk}$.

An example wherein the ICM model correctly estimates the mEE can be seen in Figure 12. The blue line represents the measured mEE $\hat{y}$ from the breath-by-breath respiratory response. The orange line displays the first-order model curve $\bar{y}$ with the subject-specific time constant of $\tau$ = 41. In this case there is a good fit between $\hat{y}$ and $\bar{y}$ and $R^2$ = .82. Moreover, it can be seen that the model outcome $\bar{\dot{E}}$ (i.e. end value of the model curve $\bar{y}$), represented by the purple line, lies within the individual $95\%CI_{meas}$ control boundaries (i.e. green dotted lines). When this is the case, the instantaneous energy expenditure estimation was considered to be correct and similar to the statistically determined steady-state $mEE_{walk}$ control value (i.e. green line, lying under the purple line). In this case the estimation had an absolute error of 0.01 [J·kg$^{-1}$· min$^{-1}$].

Figure 13 displays the single trial where the estimation is considered to be incorrect (i.e. $\bar{\dot{E}}$ does not lie within the $95\%CI_{meas}$ boundaries) with the full conventional measurement duration. In this case, the subject-specific time constant was $\tau$ = 25, the goodness-of-fit $R^2$ = 0.29, and the absolute estimation error of 7.48 [J·kg$^{-1}$· min$^{-1}$].

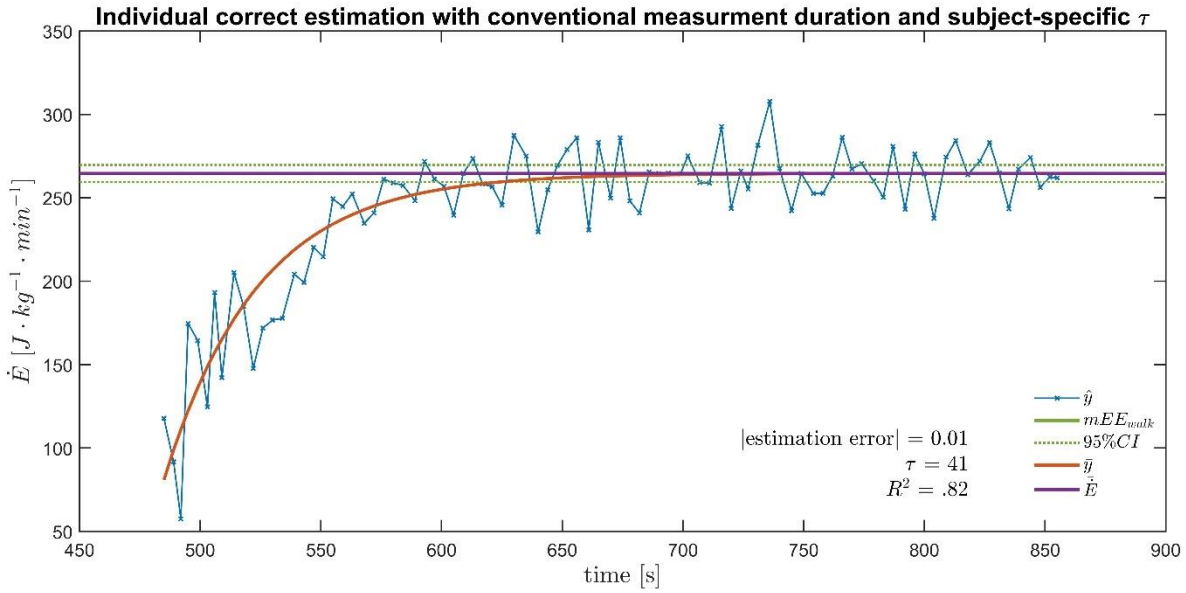



**Figure 12**. An example of a single trial wherein the ICM model correctly describes the respiratory response $\hat{y}$ of a patient with an adult neuromuscular disease. Symbols: $\hat{y}$ = respiratory response; $\bar{y}$ = model curve; $mEE_{walk}$ = steady state control value, $95\%CI_{meas}$ = individually control boundaries, $\bar{\dot{E}}$ = estimation of the instantaneous energy expenditure.

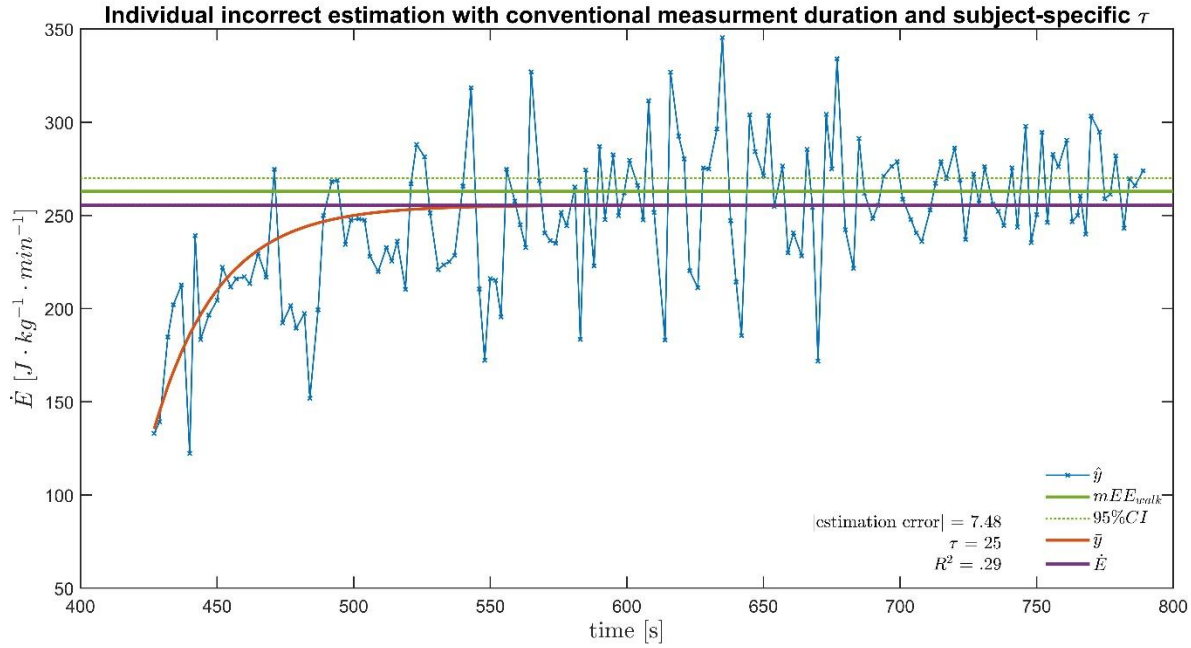



**Figure 13**. The example of the single trial wherein the ICM model does not correctly describes the respiratory response $\hat{y}$ of an adult Multiple Sclerosis patient. Symbols: $\hat{y}$ = respiratory response; $\bar{y}$ = model curve; $mEE_{walk}$ = steady state control value, $95\%CI_{meas}$ = individually control boundaries, $\bar{\dot{E}}$ = estimation of the instantaneous energy expenditure.

*3.3.1.2. Equivalence of average $\bar{\dot{E}}$ and $mEE_{walk}$*

The $mEE_{walk}$ values measured from the breath-by-breath steady windows, $D(28)$ = .16, $p$ = .06, do not deviate significantly from normal and the variances are equal among the groups, $F(2,23)$ = 2.50, $p$ = .10. Also, the ICM model outcome $\bar{\dot{E}}$ values, $D(28)$ = .14, $p$ = .14, do not deviate from normal and the variances were equal among the groups, $F(2,23)$ = 2.70, $p$ = .09.

The one-sample Equivalence Test, showed that $\bar{\dot{E}}$ ($M$ = 298.45, $SE$ = 14.05) on a group level is equivalent to the steady-state $mEE_{walk}$ ($M$ = 295.94, $SE$ = 13.56) control values, since the mean difference of 2.52 and the confidence interval of the difference $95\%CI$ [ -26.31 31.35] lie within the defined equivalence margin of $\pm\,\Delta_E$ = 38.47 [J·kg$^{-1}$· min$^{-1}$].

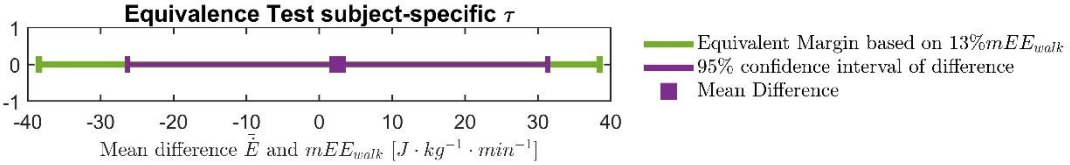



**Figure 14**. Results of the one sample Equivalence Test, illustrating that the mean difference and 95% confidence interval of the difference lie within the equivalence region [-38.47 38.47] that is based on a clinical relevant difference of 13% of the average $mEE_{walk}$.

### 3.3.2. Evaluation of the required measurement duration

*3.3.2.1. Estimation-error-measurement-duration relationship*

The individual estimation-error-measurement-duration relationship of the total population for the subject-specific model can be found in Figure 15. The group mean and 95% confidence interval relationship for the subject-specific model can be found in Figure 16. These figures show that after a measure duration of 2 minutes the mean estimation error of the subject-specific model is 3.9% on a group level. This is slightly smaller to the estimation error of 4.3% that Zhang et al. (2017) found in healthy adults using a general model. From Figure 16 it can be found that, based on the average group results, the measurement duration $\Delta t_{meas}$ can be can be reduced to a duration of 02m:33s by

accepting the 4%UBCI error level. Table 2 gives an overview of the average estimation-error-measurement-duration relationship identified among the groups in the population. An illustration of these relationships for the separate adult patient (i.e. MS and ND, $n$ = 23) and typically developing children ($n$ = 4) groups can be found in Appendix 9.2.

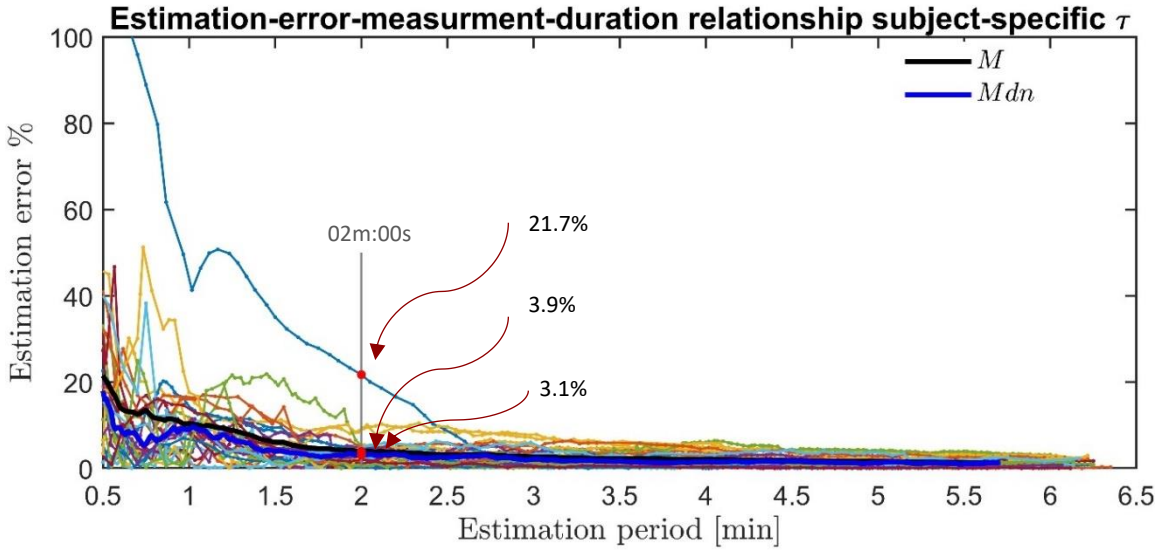



**Figure 15**. Illustration the individual, mean, and median estimation-error-measurement-duration relationship of the model using a subject-specific time constant of the total population ($n$ = 28). The group mean ($M$) and median ($Mdn$) are illustrated with the thick black and blue line, respectively. At a measurement duration of 2 minutes, the mean $M$ = 4.0%, the $Mdn$ = 3.1%, and the maximum error 21.7% (i.e. patient with Multiple Sclerosis). *Note:* illustration is inspired on Figure 11 of Zhang et al. (2017).

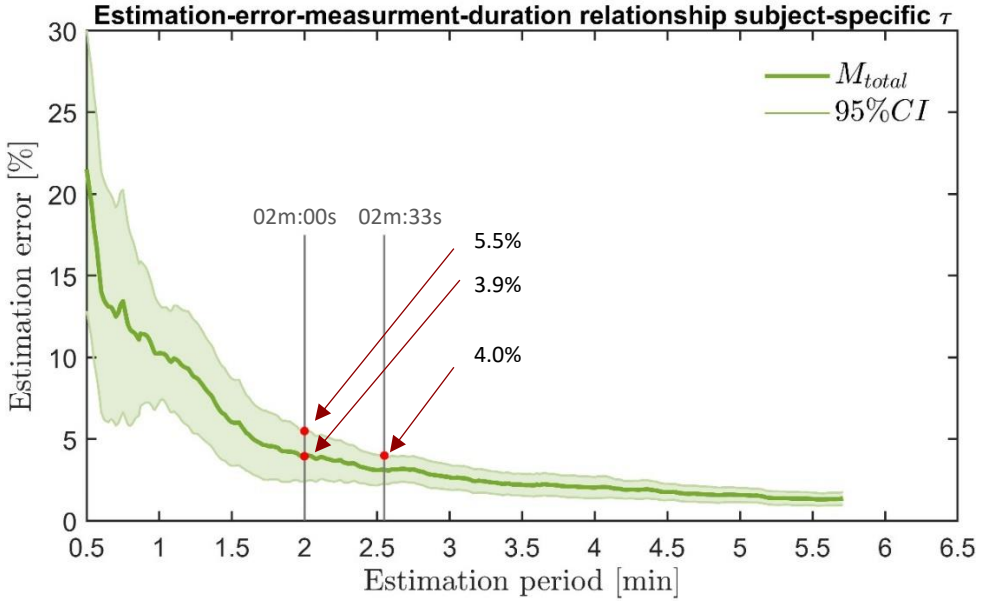



**Figure 16**. Estimation-error-measurement-duration relationship of the total population ($n$ = 28). At 2 minutes of measurement there is an average error of 3.9%, but a 5.5% error for the upper 95% confidence interval boundary. Before a 4% error level is reached in 95% of the measurement should continue 2m33s.

**Table 2. Mean estimation-error-measurement-duration relationship subject-specific model**

| | | *Measurement duration [mm:ss]* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Estimation error [%] | | *00:30* | *01:00* | *01:30* | *02:00* | *02:30* | *03:00* | *00:30* | *04:00* | *04:30* | *05:00* | *05:30* | *06:00* |
| *total* | $n$ = 28 | 21.5 | 10.3 | 6.0 | 3.9 | 3.1 | 2.7 | 2.2 | 2.0 | 1.8 | 1.6 | 1.3 | |
| *adult* | $n$ = 23 | 21.1 | 10.7 | 6.2 | 4.0 | 3.2 | 2.7 | 2.2 | 2.0 | 1.7 | 1.5 | 1.3 | 21.1 |
| ND | $n$ = 14 | 16.8 | 10.3 | 4.8 | 3.1 | 2.8 | 2.7 | 2.3 | 2.0 | 1.7 | 1.5 | 1.2 | 1.3 |
| MS | $n$ = 8 | 28.5 | 11.4 | 8.2 | 5.5 | 3.9 | 2.7 | 2.2 | 2.1 | 1.7 | 1.6 | 1.4 | |
| *child* | $n$ = 5 | 23.4 | 8.4 | 5.3 | 3.6 | 2.6 | 2.6 | 2.3 | 2.3 | 2.2 | 2.0 | 1.7 | |
| TD | $n$ = 4 | 22.4 | 8.2 | 5.8 | 4.0 | 2.7 | 2.8 | 2.7 | 2.9 | 2.6 | 2.4 | 2.0 | 1.9 |

Abbreviations: $n$ = number of participants; TD = Typical Developing children; MS = Multiple Sclerosis; ND = a variety of neuromuscular diseases. The shaded area is the first measurement duration wherein the mean %error < 4%. The single results for the healthy adult and child with Cerebral Palsy are left out. Shaded area represents the measurement duration where upon all group errors are below 4%.

*3.3.2.2. Evaluation of the reduced measurement duration on individual correctness*

When the identified reduced measurement relationship found for the 4%UBCI acceptance level of $\Delta t_{meas}$ of 2m:33s for the subject-specific model was used, 16 out of 28 (57%) estimations of the instantaneous energy expenditure could be considered as individually correct (i.e. $\bar{\dot{E}}$ lie within the individually $95\%CI_{meas}$ control boundaries). Which the present study defines as correct predictions of the steady-state energy expenditure $mEE_{walk}$. This implies that 11 estimations that were considered to be correct on an individual basis with the full conventional measurement duration are now considered as being incorrect using the shortened identified measurement duration (4%UBCI $\Delta t_{meas}$ = 2m:33s). Figure 17 shows an example of an estimation wherein the estimation becomes incorrect using the reduced measurement duration.

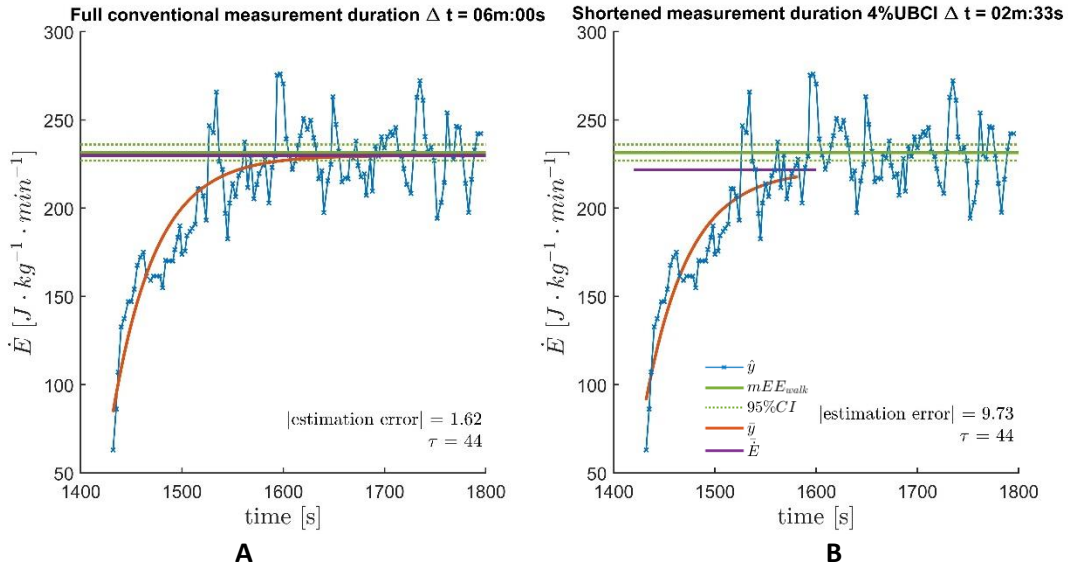



**Figure 17**. An example of a trial (in this case the healthy adult) wherein the subject-specific model A) correctly estimates the instantaneous energy expenditure using the full conventional measurement duration on an individual level. But where the subject-specific model makes B) an individually incorrect estimation using the identified reduced measurement duration of 02m33s. Symbols: $\hat{y}$ = respiratory response; $\bar{y}$ = model curve; $mEE_{walk}$ = steady state control value, $95\%CI_{meas}$ = individually control boundaries, $\bar{\dot{E}}$ = estimation of the instantaneous energy expenditure.

Since the identified reduced measurement duration resulted in a rather low rate of individual correctness (i.e. 16 out of 28, 57%) an additional acceptance level of 2% for the upper bound of the confidence interval (2%UBCI) was inspected. This resulted in, 26 out of 28 (93%) individual correct estimations, but to the drawback of a required measurement duration of 2%UBCI $\Delta t_{meas}$ = 5m04s.

*3.4. Secondary outcome measures*

3.4.1. Goodness-of-the-fit model curve $\bar{y}$

The $R^2$ values between the model and the raw signal ranged widely from .13 to .89 ($M$ = .54, $SE$ = .04). Figure 18 shows an example of a good, medium and low $R^2$ value. Low $R^2$ values could be the result of elevated noise due to a high inter-breath variability (clearly visible in Figure 18C), or could result from deviant dynamic behavior in the metabolic response (slightly visible in Figure 18C). On average, the $R^2$ values for adults ($M$ = .58, $SE$ = .04), were higher compared to the values for the children ($M$ = .34, $SE$ = .08) among the population. This difference -.24 was significant, $t$ = -2.54, $p$ = .01, $d$ = -.48.

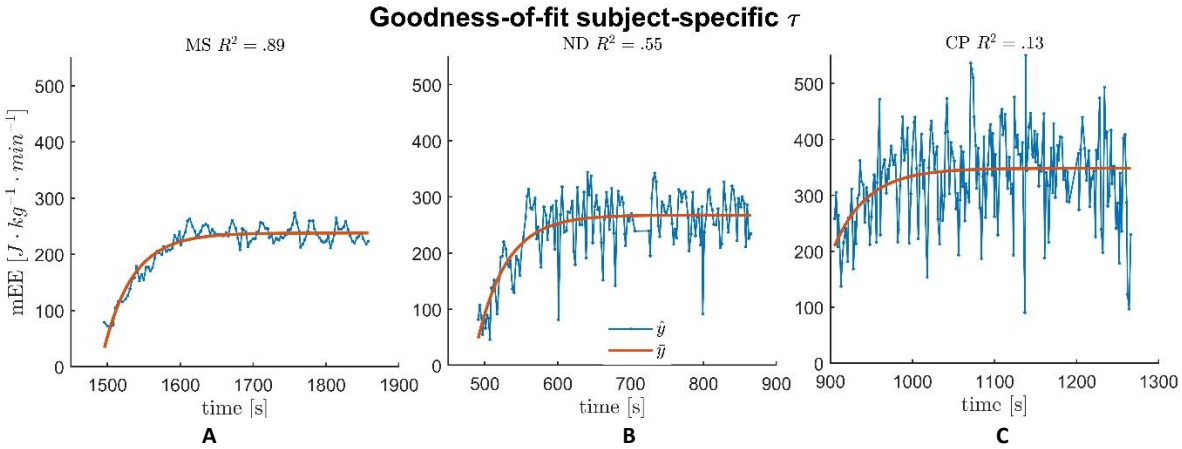



**Figure 18**. Illustration of a good, medium and bad goodness-of-fit $R^2$values of A) an adult patient with Multiple Sclerosis $R^2$=.89, B) an adult patient with a neuromuscular disease $R^2$=.55, and C) a child with Cerebral Palsy $R^2$= 0.13. The blue line represents the measured breath-by-breath metabolic energy expenditure (mEE) response $\hat{y}$ . The red line represents the identified model curve $\bar{y}$.

3.4.2. Subject-specific time constant

The subject-specific time constant $\tau$ varied widely from 23.2s to 68.9s among the total population. The identified $\tau$ did not deviate significantly from normal, $D(28)$ = .14, $p$ = .12, but the variances among the different groups in the included population were found to be not equal, $F(2,23)$ = 3.87, $p$ = .04. According to the non-parametric Independent-Samples Kruskal-Wallis test, no statistical difference was found among the different groups, $H(4)$ = 2.24, $p$ = .69, see Figure 19. Moreover, the parametric one-sample T-Test found no statistical difference between the average subject-specific time constant identified for the total population ($M$ = 41.5, $SE$ = 2.4) and the general time constant reported by Selinger and Donelan (2014) for ($n$ = 10) healthy adults ($M$ = 41.8, SE = 3.8), $t(27)$ = -.12, $p$ = .90, $d$ = -.02.

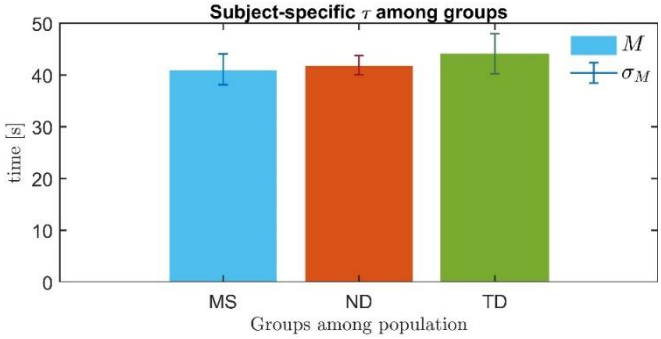


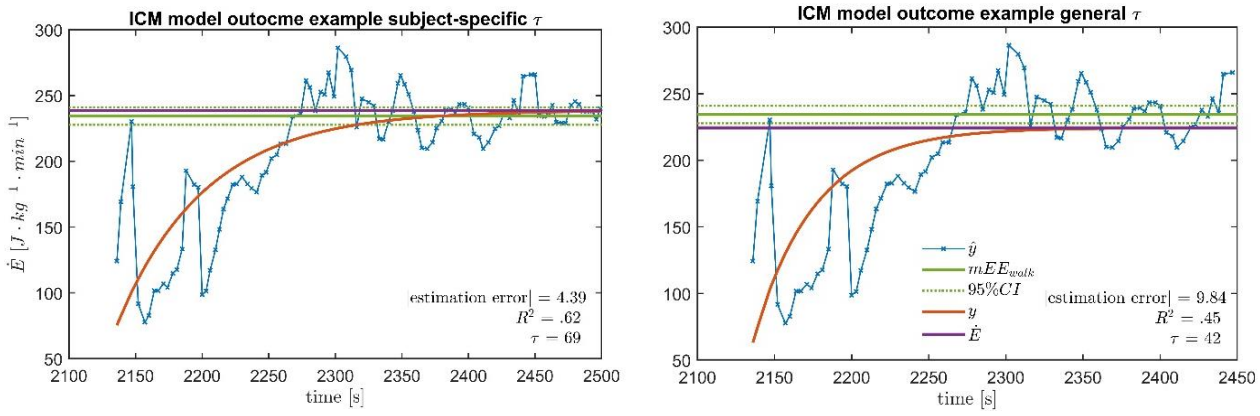

**Figure 19**. Average subject-specific time constant $\tau$ among groups with error bars representing the standard error of the mean.

### 3.4.1. Effect of general $\tau$ on the correctness of the ICM model outcome $\bar{\dot{E}}$

Since no difference among the groups in the subject-specific $\tau$ is found and the identified average $\tau$ of the population was similar to the general $\tau$ reported for healthy adults in the human-in-the-loop optimization techniques (Selinger & Donelan, 2014; Kim et al., 2017; Zhang et al., 2017; Ding et al., 2018), the present study accepted the average value of $\tau = 42$s to explore the effect of the general model on the estimation errors and required measurement duration.

#### *3.4.1.1. Individual correctness of $\bar{\dot{E}}$*

Using a general time constant a total of 25 out of 28 estimations (89%), can be considered as individually correct and $\bar{\dot{E}}$ lie within the $95\%CI_{meas}$. See Figure 20 for an example wherein the model outcome is considered as being correct with the subject-specific model but incorrect with the general model. As can be seen from the figures, using a general model increases the absolute estimation error and decreases the goodness-of-fit $R^2$ value.




**A** **B**

**Figure 20**. An example of the difference in model outcome in a patient with multiple sclerosis for the A) subjects-specific model outcome ($\tau$ = 68.7s) and B) the model outcome when a general time constant ($\tau$ =42s) is used. Symbols: $\hat{y}$ = respiratory response; $\bar{y}$ = model curve; $mEE_{walk}$ = steady state control value, $95\%CI_{meas}$ = individually control boundaries, $\bar{\dot{E}}$ = estimation of the instantaneous energy expenditure.

#### *3.4.1.2. Equivalence of average $\bar{\dot{E}}$ and $mEE_{walk}$*

The one-sample Equivalence Test showed that the general model outcome $\bar{\dot{E}}$ ($M$ = 297.39, $SE$ = 13.54) was equivalent to the steady-state $mEE_{walk}$ ($M$ = 295.94, $SE$ = 13.56) control values on a group level. Since the mean difference, of 1.45 and the $95\%CI$ [-26.34 29.24] lie within the defined equivalence margin of $\pm\,\Delta_E$ = 38.47 [J·min$^{-1}$·kg$^{-1}$].

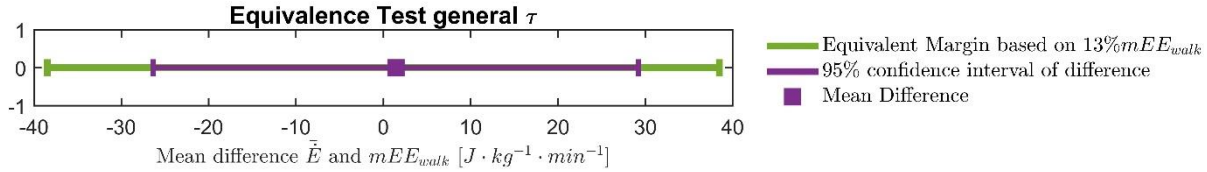



**Figure 21**. Results of the One Sample Equivalence Test for the general model outcome, illustrating that the mean difference and 95% confidence interval of the difference lie within the equivalence region [-38.47 38.47] that was based on the clinical relevant difference of 13% of the $mEE_{walk}$ (Buckon et al., 2004).

3.4.2. Effect of general $\tau$ on the estimation-error-measurement-duration relationship

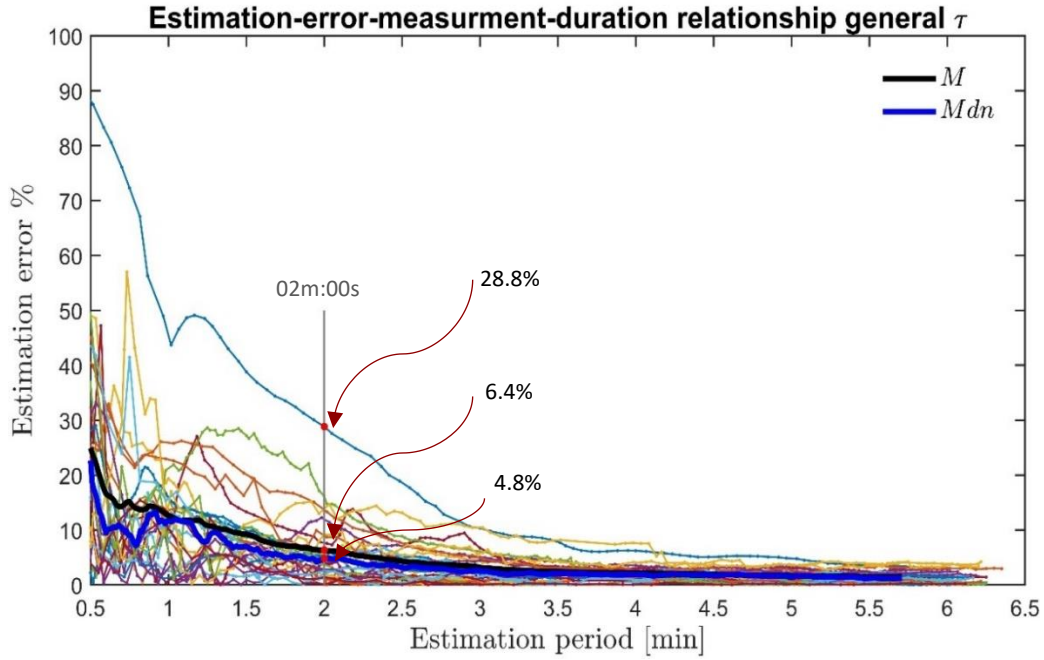



**Figure 22**. Illustration of the individual, mean, and median estimation-error-measurement-duration relationship of the model using a general time constant of 42s for the total population ($n$ = 28). The group mean ($M$) and median ($Mdn$) are shown with the thick black and blue line, respectively. At a measurement duration of 2 minutes, the mean $M$ = 6.4%, the $Mdn$ = 4.8%, and maximum error is 28.8% (i.e. patient with Multiple Sclerosis). *Note:* illustration is inspired on Figure 11 of Zhang et al. (2017).

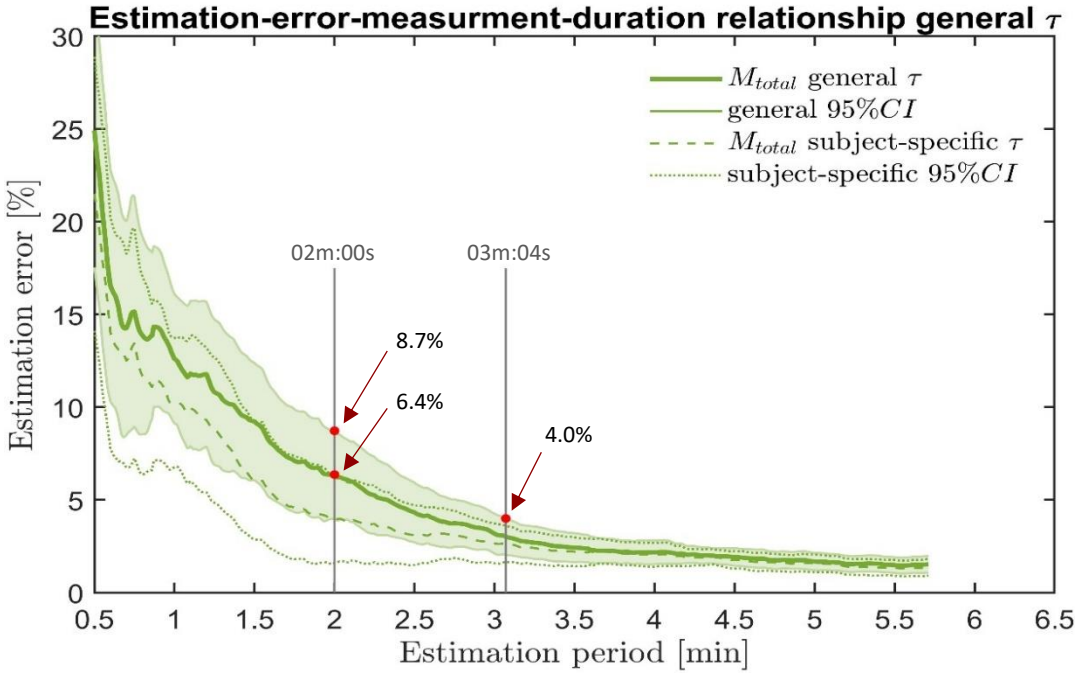



**Figure 23**. The mean and 95% confidence interval of the estimation-error-measurement-duration relationship of the total population ($n$=28) using a general time constant of 42s are represented by the solid lines and shaded area. The dotted lines represent the results of the subject-specific model as was previously shown in Figure 16. The difference between the two models becomes smaller when the measurement duration increases.

These figures show that after a measure duration of 2 minutes the mean estimation error of the general model is 6.4% on a group level. This is somewhat larger than the estimation error for the 3.9%

of the subject-specific model and the 4.3% that Zhang et al. (2017) found in healthy adults using a general model ($\tau$ = 42s). The estimation-error-measurement-duration relationship of the general model is elevated compared to the subject-specific model, but this difference attenuates for a longer measurement duration. From the figures it can also be seen that the identified measurement duration for the general model is 4%UBCI $\Delta t_{meas}$ = 3m04s. When using this identified measurement duration, a total of 16 out of 28 estimations (57%) of the instantaneous energy expenditure $\bar{\dot{E}}$ lie within the individually $95\%CI_{meas}$ boundaries. The present study considered this as individually correct estimations of the steady-state energy expenditure $mEE_{walk}$. The general 2%UBCI acceptance level resulted an individually correctness of 93% (i.e. 26 out of 28) and a required measurement duration of $\Delta t_{meas}$ = 5m11s. What can be noticed for the general model, is that 1 trial is considered to be individually correct with the reduced measurement duration of 2%UBCI $\Delta t_{meas}$ = 5m11s, while it was considered being incorrect with the full conventional measurement duration. This is probably due to the small distortion of the metabolic response dynamics in the last minute of this single trial, see Figure 24. As can be seen from the picture, the estimation of the instantaneous $\bar{\dot{E}}$ (i.e. purple line) lie just outside the individually determined $95\%CI_{meas}$ boundaries for the full conventional measurement duration.

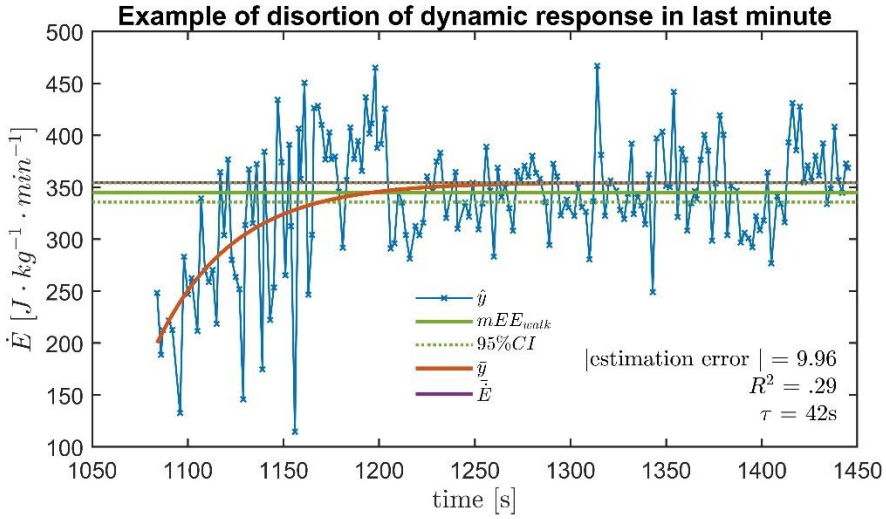



**Figure 24**. Example of the single trial that is considered to be individually incorrect with the full conventional measurement duration using a general model ($\tau$ = 24), while considered individually correct for a reduced measurement duration of 2%UBCI $\Delta t_{meas}$ = 5m11. Symbols: $\hat{y}$ = respiratory response; $\bar{y}$ = model curve; $mEE_{walk}$ = steady state control value, $95\%CI_{meas}$ = individually control boundaries, $\bar{\dot{E}}$ = estimation of the instantaneous energy expenditure.

An overview of the results for the subject-specific and general time model for the full conventional measurement duration can be found in Table 3 and the results for the reduced measurement duration in Table 4.

**Table 3. Overview of the results of subject-specific and general model for the total conventional measurement duration**

| | | | STEADY-STATE OUTCOME | | | | SUBJECT-SPECIFIC MODEL OUTCOME | | | | | | | | GENERAL MODEL OUTCOME ($\tau = 42s$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $n$ | $mEE_{walk}$ | | 95% $CI_{meas}$ | | $\tau$ | | $\dot{\bar{E}}$ | | estimation error | | $R^2$ | | $\dot{\bar{E}}$ | | estimation error | | $R^2$ | |
| *heathy* | *patient* | | *J·kg⁻¹·min⁻¹* | | *J·kg⁻¹·min⁻¹* | | s | | *J·kg⁻¹·min⁻¹* | | *%* | | *ratio* | | *J·kg⁻¹·min⁻¹* | | *%* | | *ratio* | |
| | | 22/5 | *M* | *SD* | $M_{low}$ | $M_{up}$ | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| *Total* | | | 295.9 | 71.7 | 286.6 | 305.3 | 41.5 | 12.9 | 298.5 | 74.4 | 1.4 | 0.9 | 0.54 | 0.21 | 297.4 | 71.7 | 1.4 | 1.1 | 0.53 | 0.21 |
| *Adult* | | 25 | 273.6 | 46.5 | 265.3 | 281.9 | 41.7 | 11.6 | 275.4 | 47.3 | 1.2 | 0.9 | 0.58 | 0.20 | 274.9 | 47.6 | 1.4 | 1.1 | 0.57 | 0.20 |
| HA□ | | 1 | 231.5 | - | 227.0 | 236.0 | 44.3 | - | 229.9 | - | 0.7 | - | 0.75 | 0.00 | 229.2 | - | 1.0 | - | 0.74 | - |
| | MS | 8 | 269.8 | 32.1 | 262.8 | 276.8 | 41.1 | 15.6 | 271.0 | 31.6 | 1.3 | 0.9 | 0.58 | 0.22 | 270.3 | 33.5 | 1.5 | 1.0 | 0.56 | 0.22 |
| | ND | 14 | 278.8 | 54.1 | 269.5 | 288.1 | 41.9 | 9.9 | 281.2 | 55.1 | 1.2 | 0.9 | 0.57 | 0.19 | 280.9 | 54.9 | 1.5 | 1.2 | 0.56 | 0.19 |
| *Child* | | 7 | 398.7 | 82.3 | 384.2 | 413.3 | 40.5 | 19.4 | 404.3 | 89.3 | 1.9 | 0.8 | 0.34 | 0.18 | 400.7 | 77.2 | 1.5 | 1.0 | 0.33 | 0.17 |
| TD | | 1 | 410.8 | 89.8 | 396.3 | 425.2 | 44.1 | 20.4 | 419.4 | 95.5 | 2.0 | 0.9 | 0.13 | 0.00 | 413.8 | 82.5 | 1.7 | 1.1 | 0.13 | - |
| | CP□ | 4 | 350.5 | - | 335.6 | 365.5 | 25.8 | - | 344.1 | - | 1.8 | - | 0.39 | 0.15 | 348.3 | - | 0.6 | - | 0.38 | 0.14 |

□ Single measured value is presented since no mean was possible. All data are presented by the mean (*M*) and standard deviation (*SD*) unless other denoted. Abbreviations: HA = Healthy Adults; TD = Typical Developing children; MS = Multiple Sclerosis; ND = a variety of Neuromuscular Diseases; CP = Cerebral Palsy; $n$ = number of participants; $mEE_{walk}$= steady-state metabolic energy expenditure control values; $\dot{\bar{E}}$ = the estimated instantaneous metabolic energy expenditure model outcome; $R^2$= goodness-of-the-fit parameter; $\tau$ = subject-specific time constant. All data is computed according the equation of Garby and Astrub (1987) which should be considered in case of comparison with results of different studies.

**Table 4. Overview of the results of the subject-specific and general model for the identified reduced measurement duration**

| | | | SUBJECT-SPECIFIC MODEL OUTCOME with 4%UBCI $\Delta t_{meas}$ = 2m33s | | | | | | GENERAL MODEL OUTCOME ($\tau = 42s$) with 4%UBCI $\Delta t_{meas}$ = 3m04s | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $n$ | $\dot{\bar{E}}$ | | estimation error | | $R^2$ | | $\dot{\bar{E}}$ | | estimation error | | $R^2$ | |
| | | | *J·kg⁻¹·min⁻¹* | | *%* | | *ratio* | | *J·kg⁻¹·min⁻¹* | | *%* | | *ratio* | |
| *heathy* | *patient* | 22/5 | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| *Total* | | | 298.5 | 74.4 | 1.37 | 0.90 | 0.54 | 0.21 | 297.4 | 71.7 | 1.45 | 1.05 | 0.53 | 0.21 |
| *Adult* | | 25 | 275.4 | 47.3 | 1.25 | 0.89 | 0.58 | 0.20 | 274.9 | 47.6 | 1.44 | 1.08 | 0.57 | 0.20 |
| HA□ | | 1 | 229.9 | - | 0.70 | - | 0.75 | - | 229.2 | - | 1.00 | - | 0.74 | - |
| | MS | 8 | 271.0 | 31.6 | 1.31 | 0.91 | 0.58 | 0.22 | 270.3 | 33.5 | 1.47 | 1.01 | 0.56 | 0.22 |
| | ND | 14 | 281.2 | 55.1 | 1.25 | 0.94 | 0.57 | 0.19 | 280.9 | 54.9 | 1.45 | 1.18 | 0.56 | 0.19 |
| *Child* | | 7 | 404.3 | 89.3 | 1.93 | 0.78 | 0.34 | 0.18 | 400.7 | 77.2 | 1.47 | 1.04 | 0.33 | 0.17 |
| TD | | 4 | 419.4 | 95.5 | 1.95 | 0.90 | 0.39 | 0.15 | 413.8 | 82.5 | 1.68 | 1.08 | 0.38 | 0.14 |
| | CP□ | 1 | 344.1 | - | 1.85 | - | 0.13 | - | 348.3 | - | 0.63 | - | 0.13 | - |

□ Single measured value is presented since no mean is possible with ($n$ = 1). All data are presented by the mean (*M*) and standard deviation (SD) unless other denoted. Abbreviations: HA = Healthy Adults; TD = Typical Developing children; MS = Multiple Sclerosis; ND = a variety of Neuromuscular Diseases; CP = Cerebral Palsy; $n$ = number of participants; $mEE_{walk}$= steady-state metabolic energy expenditure control values; $\dot{\bar{E}}$ = the estimated instantaneous metabolic energy expenditure model outcome; $R^2$= goodness-of-the-fit parameter; $\tau$ = subject-specific time constant. All data is computed according the equation of Garby and Astrub (1987) which should be considered in case of comparison with results of different studies.

# 4. Discussion

The aim of the present study was to investigate whether the ICM model could be applied and reduce the measurement duration of the metabolic gas analysis in a variety of child and adult patients with neurology and neuromuscular diseases with subsequent gait impairments and healthy controls. The ICM model originates from upcoming human-in-the-loop device optimization techniques and make use of a first-order differential equation with an additional time constant. It was expected that the respiratory dynamics of the patients is similar to healthy subjects. It was therefore hypothesized that the ICM model could correctly estimate the instantaneous energy expenditure of the included population. It was expected that the subject-specific ICM model could reduce the conventionally measurement duration of 6 minutes, to approximately 2-3 minutes, similar as in healthy adult subjects in the human-in-the-loop techniques. Additionally, the use of the general ICM model with a fixed time constant was evaluated for the correctness of the estimations of the and the required measurement duration. It was expected that a general model would result higher estimation errors. Leading to a longer required measurement duration. Nevertheless, it was expected that the identified measurement duration for the general model was still less than the conventional duration of 6 minutes.

### *4.1. Interpretation of key results*

The results indicate that the subject-specific ICM model can correctly estimates the instantaneous energy expenditure using the full conventional measurement duration. On an individual level, 96% of the estimations were considered as being correct. This implies that in only 1 out of the 28 subjects, the model outcome did not satisfy the individually defined $95\%CI_{meas}$ control boundaries. Determined from the steady time windows in the breath-by-breath mEE. Furthermore, on a group level, the analysis confirms that the average estimation of the instantaneous energy expenditure of the population is equivalent to the averaged steady-state metabolism.

In line with the hypothesis, the identified measurement duration ($\Delta t_{meas}$ = 2m33s) for the subject-specific model, using a 4%UBCI acceptance level, is similar to the 2-3 minutes used in the human-in-the-loop techniques. However, on an individual level, this reduced measurement duration resulted in a rather low individually correctness of only 57%. This implies that only 16 out of 28 estimations satisfied the individually defined $95\%CI_{meas}$ control boundaries.

No differences were found in the identified subject-specific time constants among the groups in the included population. This result contradicts the claim that children have faster oxygen kinetics (Fawkner et al., 2002; Armstrong & Barker, 2009) leading to smaller subject-specific time constants. However, this result deserves caution since the size of the groups among the population were rather small and unequally sampled. No comparison was possible between the HA and CP groups since only one participant per group was included. The results indicate that the average identified $\tau$ value do not differ from the $\tau$ reported in literature for healthy adults (Selinger & Donelan, 2014). Therefore, the present study has taken this average value ($\tau$ = 42s) as fixed time constant for the general model.

In line with the expectations, the general model results in increased estimation errors compared to the subject-specific model on both an individual and group level. On the individual level, the general model reduces the individually correctness of the estimations to 89% (i.e. 25 out of 28) for the full conventional measurement duration. On group level, the results show that the estimation-error-measurement-duration relationship is elevated compared to the subject-specific model. This difference attenuates over the increasing measurement duration. Consequently, the identified average required measurement duration for the general model is longer (4%UBCI $\Delta t_{meas}$ = 3m04s) than for the subject-specific model. But still, when an acceptance level of 4%UBCI is handled, the measurement duration is similar to the expected 2-3 minutes used in human-in-the loop techniques. When using the identified reduced measurement duration, the general model resulted in a rather low individually correctness rate of 57%, similar as with the subject-specific model.

These low individually correctness rates, raise the question whether the use of a 4% estimation error acceptance level adopted from Zhang et al. (2017) is appropriate to evaluate the potential use of the ICM model on an individual level. Therefore, also a 2%UBCI acceptance level was alternatively explored and resulted in an improved individual correctness rate of 93% (i.e. 26 out of 28) for both the subject-specific and general model. But to the drawback of the increased required measurement durations of $\Delta t_{meas}$ = 5m04s and $\Delta t_{meas}$ = 5m11s for the subject-specific and general model, respectively. Which are not much shorter than the conventional 6 minutes.

The goodness-of-fit $R^2$ values showed a large variation (.13 to .89) among the subjects. Selinger and Donelan (2014) found a similar range (.26 to .96) in $R^2$ values in healthy adults. On average, the $R^2$ values showed moderate values within adults and low values among the children of the included population. This difference can be explained by the fact that children have a higher inter-breath-variability (Armstrong & Barker, 2009) and does not necessarily mean that the model does not correctly fit the respiratory response.

#### 4.1.1. Can the ICM model reduce the measurement duration?

The results suggest that the ICM model can make individually correct estimations of the instantaneous energy expenditure, when using the full measurement duration. Handling the same error acceptance level on a group level as used within the human-in-the-loop techniques, the results show that the measurement duration can be shortened to a similar duration of 2-3 minutes.

However, the present study developed a method to evaluate the effect of reduced measurement durations on the individual correctness of the estimations. As far as known, this is something the human-in-the-loop techniques did not explored before. This individual correctness is of great importance when evaluating the potential clinical application of the ICM model for rehabilitative purposes on an individual level. This method showed that the individual correctness was reduced significantly by using the identified measurement durations.

Nevertheless, the ICM model seems promising since the results are highly comparable to the results of healthy adults in the human-in-the-loop techniques. Further improvement of the measurement protocols might improve these results even more. However, it is important to keep in mind that good results on a group level does not implicitly assure that the estimations are individually correct. This also holds for the current human-in-the-loop techniques in healthy adults.

With respect to the clinical application of the ICM model, it is recommended to consider what balance should be stricken between the estimation errors and measurement duration for each specific application. It might be valid to use the ICM technique for the purpose of within-subject walking device evaluations, where many evaluations need to be assessed. However, for a single evaluation of the gait efficiency on individual level, considering the current ICM model, it might be more appropriate to use the conventional technique. Especially, when the results are used for comparisons among subjects or evaluations over time. Future research is needed to investigate whether improved measurement protocols could improve the ICM model outcome, becoming a potential accurate replacement of the conventional techniques.

### *4.2. Limitations*

#### 4.2.1. Small and unequal sample size

The main limitation of the present study is the small sample size. Although a lot of trials ($n$ = 482) were collected, only a small number of trial ($n$ = 28) were included due to strict inclusion criteria that reduces the generalizability (see Appendix 9.3 for the considerations). The present study had to deal with the (in)completeness of the post hoc data and cope with unforeseen data collecting challenges (see Appendix 9.4 for a description). The heterogeneity of the included population was high, including different pathologies, walking conditions (i.e. shoes, AFO's, helping aids), and protocols (i.e. treadmill, over-ground track), making it difficult to draw firm conclusions. However, the present study is the first that explores the clinical usability of the ICM model in patients and typically developing children and due to the strict inclusion criteria, a clean dataset was used for analysis.

#### 4.2.1. Clinically relevant difference

The individually determined $95\%CI_{meas}$ boundaries used to define whether the estimations were considered to be individually correct is somewhat stricter than the golden standard clinically relevant difference used for energy expenditure. This might have distorted the interpretation of the (in)correct of the model outcome. However, caution was taken to assure that the correct estimations were truly correct and reduce the false positives. Besides, the regular clinical relevant difference takes into account day-to-day variability while in this case the test and control values originated from a single measured signal ($mEE_{walk}$ *vs.* $\bar{\dot{E}}$), a strict acceptance level seems therefore appropriate.

#### 4.2.2. Step change in workload

A limitation of the current ICM model is that it only models the response to a step change in workload. After collecting and exploring the data it became clear that some trials were no clear step changes in workload or did not have consistent walking speeds. For the current analysis, these trials were excluded from analysis. It was beyond the scope of the present study, but it might be possible to adapt the input signal to the external workload of the subjects to develop a more variable model response where upon the data can be fit, increasing the applicability of the model. The post hoc data did not contain enough information to retrospectively design the external workload input function and would not be usable for such an approach.

### *4.3. Recommendations*

To better understand the implications of these results, future studies with a larger sample size are required to investigate whether there are differences in subject-specific time constants among the groups of the population. It could be investigated whether the use of a disease-specific time constant could improve the results of the general model, without the impractical need to identify the subject-specific time constant in advance. For future studies, it is recommended to collect own data instead of post hoc data so that the research set-up can be improved to the needs of the ICM technique and the required information can be collected. For example, adaptations in the research set-up can assure that subjects start from a rest metabolism by performing the rest measurement directly prior the walking trial at the same location of the walking track/treadmill (which was not the case in the Amsterdam UMC). Moreover, accurate measures of the walking speed improves a valid in- or exclusion of trials based on the consistency in walking speed or can be used to identify a variable external workload used to adapt the ICM model response, as explained in 4.2.2. This seems possible with the GRAIL system. However, treadmill belt speed collection was cumbersome, and the signals contained large offsets. In addition, for clinical reasons, it is often preferred to walk over-ground in patients with severe gait impairments and additional helping aids due to a reduced stability.

Concerning the maintenance of data, it is advised to improve the data management in a systematic manner (for example assemble copies of the raw data, used protocol, research outcomes, and associated report or thesis). This would make it more feasible to use post hoc data for other research objectives in the future and could save a lot of time. Increasing the efficiency of research.

The large exclusion rate raises the question whether the trials were usable for the original clinical research objective in the first place. It is a well-known fact that walking speed is an important determinant of the gait efficiency. As far as known, the rehabilitation departments of the Amsterdam UMC do not handle an objective measure to evaluate the consistency of walking speed in the clinical trials. While the average walking speed is also used for normalization of the energy expenditure. Whether this might be due to loss of attention on the walking speed outcome, or simply is a consequence of practical considerations is unknown. Nevertheless, it might be good to keep a critical attitude towards the current proceedings. This also holds for the determination of the steady-state metabolism. The current clinical procedure either prone to errors (i.e. using a fixed window wherein unforeseen act might happen such as couth, laugh, distraction or talking) or rather subjective and not well reproducible (i.e. selecting a steady window based on visual inspection). The fact that the objective method of Schwartz (2007) excluded 46 of the 202 trials based on the steadiness of the

metabolic data, raises the question whether either the method of Schwartz (2007) is too strict and not practical for clinical use, or that the current clinicals proceedings might be improved.

# 5. Conclusions

The present study explored whether a subject-specific and general ICM model could reduce the conventional measurement duration (6 minutes) of metabolic gas analyses in patients with gait disabilities and healthy controls. The ICM model make use of a first-order differential equation with an additional time constant that is fit to the respiratory response of the subjects that perform a walking test. This model is designed to makes fast estimates (2-3 minutes) of the instantaneous energy expenditure during gait and originates from upcoming human-in-the-loop techniques for wearable walking device optimization. The present study collected post hoc metabolic gas data within two rehabilitation departments of the Amsterdam UMC. The data contained a wide variety of neuromuscular and neurology diseased adult and child patients and healthy controls. After strict inclusion criteria, a total of 28 subjects were included for analysis among which 8 MS, 14 ND, 1 CP, 1 HA and 4 TD subjects. The present study evaluated the performance of the model on both a group and individual level to inspect whether the ICM model could accurately speed up the traditional metabolic gas analyses.

In line with the expectations, the subject-specific model improved the estimation-error-measurement-duration relationship compared to a general model. However, the present study did not identify significant differences in the subject-specific time constants among the groups in the included population. Due to the small and unequal sized samples these results should be treated with caution. From results on a group level, it can be concluded that both the subject-specific and general ICM model can reduce the measurement duration based on an average 4% UBCI acceptance level for the total population. These results are in line with the results of healthy adults in the human-in-the-loop techniques. This offers prospect and opens ways for the clinical application of the human-in-the-loop optimization techniques applied for wearable walking device tuning (e.g. AFOs) in rehabilitation. However, on an individual level, the analysis showed that a reduced measurement duration resulted in elevated estimation errors in the estimations, reducing the individually correctness of the estimations. Therefore, clinicians and researchers should consider what balance they want to strike between the clinically correctness of the estimations and the feasibility of the measurement duration. At this moment the ICM model cannot be a fully accurately replacement of the conventional metabolic gas analyses. When assessing the gait efficiency for a single walking trial, it is recommended to use the conventional measure duration. However, when many evaluations need to be assessed it might be valid to use the ICM model and accept elevated estimation errors.

Although the strict inclusion criteria might have reduced the generalizability of the results, this study provides new insights into the potential clinical application of the ICM model and highlight the importance of the inspection of the individually correctness of the upcoming human-in-the-loop techniques. Future research is required to investigate whether there are differences in subject-specific time constant among the different pathologies and whether the use of a disease-specific model might improve the results of the general model without the need of identifying a subject-specific time constant in advance. It is recommendable to use strict measurement protocols to enhance subjects to start from a rest metabolism and accurately measure the consistency in walking speed. It was beyond the scope of the present study, but it would be interesting to investigate whether the ICM model could be adaptable to varying changes in workload in the future.

# 6. Abbreviations

| | |
|---|---|
| mEE | Metabolic energy expenditure |
| COM | Center of mass |
| ATP | Adenosine triphosphate |
| MS | Multiple sclerosis |
| CP | Cerebral palsy |
| PPS | Post-polio syndrome |
| ND | Neuromuscular disorders |
| AFO | Ankle-foot-orthosis |
| $m\dot{V}O_2$ | Oxygen consumption in the muscle tissues |
| $\dot{V}O_2$ | Pulmonary metabolism oxygen consumption measured at the mouth |
| $\dot{V}CO_2$ | Carbon dioxide production |
| mECS | Metabolic energy consumption |
| COT | Cost of transport |
| mEC | Metabolic energy cost |
| ICM | Instantaneous cost mapping |
| $\tau_{\dot{V}O_2}$ | Time constant for oxygen kinetics |
| RER | Respiratory exchange ratio |
| UMC | Amsterdam university medical centers |
| AMC | Amsterdam medical center hospital |
| VUmc | Free university medical center hospital |
| TD | Typically developing children |
| MWT | Minute walk test |
| $mEE_{walk}$ | Statistically determined steady-state metabolism of the rest measurement |
| $mEE_{rest}$ | Statistically determined steady-state metabolism of the walk measurement |

# 7. Acknowledgements

The present study would like to especially thank the clinicians at the Amsterdam UMC rehabilitation departments that make this study possible and took their sparsely time to help set up the research.

# 8. References

Armstrong, N., & Barker, A. R. (2009). Oxygen uptake kinetics in children and adolescents: a review. *Pediatric exercise science*, *21*(2), 130-147.

Barker, A. R., Welsman, J. R., Fulford, J., Welford, D., Williams, C. A., & Armstrong, N. (2008). Muscle phosphocreatine and pulmonary oxygen uptake kinetics in children at the onset and offset of moderate intensity exercise. *European journal of applied physiology*, *102*(6), 727-738.

Behnke, B. J., Barstow, T. J., Kindig, C. A., McDonough, P., Musch, T. I., & Poole, D. C. (2002). Dynamics of oxygen uptake following exercise onset in rat skeletal muscle. *Respiratory physiology & neurobiology*, *133*(3), 229-239.

Bertram, J. E. and Ruina, A. (2001). Multiple walking speed- frequency relations are predicted by constrained optimization. *Journal of Theoretical Biology*, *209*(4):445–453.

Bolster, E. A., Balemans, A. C., Brehm, M. A., Buizer, A. I., & Dallmeijer, A. J. (2017). Energy cost during walking in association with age and body height in children and young adults with cerebral palsy. *Gait & posture*, *54*, 119-126.

Bregman, D. J. J., Harlaar, J., Meskers, C. G. M., & De Groot, V. (2012). Spring-like Ankle Foot Orthoses reduce the energy cost of walking by taking over ankle work. *Gait & posture*, *35*(1), 148-153.

Brehm, M. A., Becher, J., & Harlaar, J. (2007). Reproducibility evaluation of gross and net walking efficiency in children with cerebral palsy. *Developmental Medicine & Child Neurology*, *49*(1), 45-48.

Brehm, M. A., Harlaar, J., & Groepenhof, H. (2004). Validation of the portable VmaxST system for oxygen-uptake measurement. *Gait & posture*, *20*(1), 67-73.

Brehm, M. A., Knol, D. L., & Harlaar, J. (2008). Methodological considerations for improving the reproducibility of walking efficiency outcomes in clinical gait studies. Gait & posture, 27(2), 196-201.

Buckon, C. E., Thomas, S. S., Jakobson-Huston, S., Moor, M., Sussman, M., & Aiona, M. (2004). Comparison of three ankle–foot orthosis configurations for children with spastic diplegia. *Developmental Medicine and Child Neurology*, *46*(9), 590-598.

Collins, S. H., Adamczyk, P. G., and Kuo, A. D. (2009). Dynamic arm swinging in human walking. *Proceedings of the Royal Society B: Biological Sciences, 276*(1673):3679–3688.

Collins, S. H., Wiggin, M. B., & Sawicki, G. S. (2015). Reducing the energy cost of human walking using an unpowered exoskeleton. *Nature*, *522*(7555), 212.

Dal, U., Erdogan, T., Resitoglu, B., & Beydagi, H. (2010). Determination of preferred walking speed on treadmill may lead to high oxygen cost on treadmill walking. *Gait & posture*, *31*(3), 366-369.

Donelan, J. M., Kram, R., and Kuo, A. D. (2001). Mechanical and metabolic determinants of the preferred step width in human walking. *Proceedings of the Royal Society B: Biological Sciences, 268*(1480):1985–1992.

Donelan, J. M., Kram, R., & Kuo, A. D. (2002). Simultaneous positive and negative external mechanical work in human walking. *Journal of biomechanics, 35*(1), 117-124.

Fawkner, S. G., Armstrong, N., Potter, C. R., & Welsman, J. R. (2002). Oxygen uptake kinetics in children and adults after the onset of moderate-intensity exercise. *Journal of sports sciences*, *20*(4), 319-326.

Felt, W., Selinger, J. C., Donelan, J. M., & Remy, C. D. (2015). " Body-In-The-Loop": Optimizing Device Parameters Using Measures of Instantaneous Energetic Cost. *PloS one*, *10*(8), e0135342.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics 4th edition*. London, United Kingdom: Sage publications Ltd

Geurtz, B., & Los, F.S. (2016). Does walking-specific resistance training change walking related motor fatigue in patients with MS? [Msc thesis, VU University, Faculty of Human Movement Sciences].

Kerkum, Y. L., Buizer, A. I., van den Noort, J. C., Becher, J. G., Harlaar, J., & Brehm, M. A. (2015). The effects of varying ankle foot orthosis stiffness on gait in children with spastic cerebral palsy who walk with excessive knee flexion. *PloS one*, *10*(11), e0142878.

Kim, M., Ding, Y., Malcolm, P., Speeckaert, J., Siviy, C. J., Walsh, C. J., & Kuindersma, S. (2017). Human-in-the-loop Bayesian optimization of wearable device parameters. *PloS one*, *12*(9), 1-15. doi:https://doi.org/10.1371/journal.pone.0184054

Kipp, S., Byrnes, W. C., & Kram, R. (2018). Calculating metabolic energy expenditure across a wide range of exercise intensities: the equation matters. *Applied Physiology, Nutrition, and Metabolism, 43*(6), 639-642.

Koller, J. R., Gates, D. H., Ferris, D. P., & Remy, C. D. (2016, June). 'Body-in-the-Loop' Optimization of Assistive Robotic Devices: A Validation Study. In *Robotics: Science and Systems* (Vol. 2016, pp. 1-10).

Koller, J. R., Gates, D. H., Ferris, D. P., and Remy, C. D. (2017). Confidence in the curve: Establishing instantaneous cost mapping techniques using bilateral ankle exoskeletons. *Journal of Applied Physiology, 122*(2), 242-252.

Kuo, A. D. (2001). A Simple Model of Bipedal Walking Pre- dicts the Preferred Speed–Step Length Relationship. *Journal of Biomechanical Engineering, 123*(3):264.

Kuo, A. D., Donelan, J. M., & Ruina, A. (2005). Energetic consequences of walking like an inverted pendulum: step-to-step transitions. *Exercise and sport sciences reviews, 33*(2), 88-97.

Lamarra, N., Whipp, B. J., Ward, S. A., & Wasserman, K. (1987). Effect of interbreath fluctuations on characterizing exercise gas exchange kinetics. *Journal of Applied Physiology*, *62*(5), 2003-2012.

Malcolm, P., Rossi, D. M., Siviy, C., Lee, S., Quinlivan, B. T., Grimmer, M., and Walsh, C. J. (2017). Continuous sweep versus discrete step protocols for studying effects of wearable robot assistance magnitude. *Journal of NeuroEngineering and Rehabilitation, 14*(1):72.

Mooney, L. M. and Herr, H. M. (2016). Biomechanical walking mechanisms underlying the metabolic reduction caused by an autonomous exoskeleton. *Journal of NeuroEngineering and Rehabilitation, 13*(1):1–12.

Plasschaert, F. S., Matthews, P. A., & Forward, M. J. (1999). Repeatability and variability of energy cost measurements. *Gait & Posture*, *1*(10), 71.

Potter, C. R., & Unnithan, V. B. (2005). Interpretation and implementation of oxygen uptake kinetics studies in children with spastic cerebral palsy. *Developmental medicine and child neurology*, *47*(5), 353-357.

Quesada, R. E., Caputo, J. M., and Collins, S. H. (2016). Increasing ankle push-off work with a powered prosthesis does not necessarily reduce metabolic rate for transtibial amputees. *Journal of Biomechanics, 49*(14):3452–3459.

Saunders, M., Inman, V., and Eberhart, H. (1953). The major determinant in normal and pathological gait. *Journal of Bones and Joint Surgery, 35*(3):543-558.

Schwartz, M. H. (2007). Protocol changes can improve the reliability of net oxygen cost data. *Gait & posture, 26*(4), 494-500.

Schwartz, M. H., Koop, S. E., Bourke, J. L., & Baker, R. (2006). A nondimensional normalization scheme for oxygen utilization data. *Gait & posture*, *24*(1), 14-22.

Selinger, J. C. and Donelan, J. M. (2014). Estimating instantaneous energetic cost during non-steady-state gait. *Journal of Applied Physiology, 117*(11):1406-1415.

Slade, P., Troutman, R., Kochenderfer, M. J., Collins, S. H., & Delp, S. L. (2019). Rapid energy expenditure estimation for ankle assisted and inclined loaded walking. *Journal of neuroengineering and rehabilitation*, *16*(1), 67.

Unnithan, V. B., Dowling, J. J., Frost, G. A. I. L., & Bar-Or, O. D. E. D. (1996). Role of cocontraction in the O2 cost of walking in children with cerebral palsy. *Medicine and science in sports and exercise*, *28*(12), 1498-1504.

Waters, R. L., & Mulroy, S. (1999). The energy expenditure of normal and pathologic gait. *Gait & posture, 9*(3), 207-231.

Waterval, N. F., Nollet, F., Harlaar, J., & Brehm, M. A. (2017). Precision orthotics: optimising ankle foot orthoses to improve gait in patients with neuromuscular diseases; protocol of the PROOF-AFO study, a prospective intervention study. *BMJ open, 7*(2), e013342.

Whipp, B. J., Ward, S. A., Lamarra, N., Davis, J. A., & Wasserman, K. (1982). Parameters of ventilatory and gas exchange dynamics during exercise. *Journal of Applied Physiology*, *52*(6), 1506-1513.

Wutzke, C. J., Sawicki, G. S., & Lewek, M. D. (2012). The influence of a unilateral fixed ankle on metabolic and mechanical demands during walking in unimpaired young adults. *Journal of biomechanics*, *45*(14), 2405-2410.

Zamparo, P., Francescato, M. P., De Luca, G., Lovati, L., & di Prampera, P. E. (1995). The energy cost of level walking in patients with hemiplegia. *Scandinavian journal of medicine & science in sports*, *5*(6), 348-352.

# 9. Appendix

## *9.1. Metabolic energy expenditure calculation methods*

As mentioned in the methods there are multiple ways to calculate the energy expenditure of gait. This section will discuss and compare the most common used calculation methods in more detail.

### 9.1.1. Different equations

A common used equation in literature (Donelan et al., 2001; Collins, Adamczyk, & Kuo, 2009; Wutzke, Sawicki, & Lewek, 2012; Malcolm et al., 2013; Mooney, & Herr, 2016; Kim et al., 2017; Zhang et al., 2017; Ding et al., 2018; Lee et al., 2018) is a simplified version of the equation of Brockway (1987), see equation (A1.).

$$mEE = 0.278 \cdot \dot{V}O_2 + 0.075 \cdot \dot{V}CO_2 \qquad (A1.)$$

Where $mEE$ is expressed in Watts and the $\dot{V}O_2$ and $\dot{V}CO_2$ the volumetric flow rates in mL·min$^{-1}$. Other studies and the clinical service in the Amsterdam UMC (Brehm et al., 2004; Kerkum et al., 2015) use the equation of Garby and Astrub (1987), see equation (A2.). Both equations (A1.) and (A2.) are based on heat combustion measurements of mixtures of oxidized carbohydrates, lipids and proteins (Brehm et al., 2004; Kipp, Byrnes, & Kram, 2018).

$$mEE = (4.960 \cdot RER + 16.040) \cdot \dot{V}O_2 \qquad (A2.)$$

Where $mEE$ is expressed in J·min$^{-1}$ and $RER$ stands for the respiratory exchange ratio that gives insight in the relative contribution of the carbohydrates and lipids that are converted into metabolic energy. The $RER$ can be calculated from the carbon dioxide production ($\dot{V}CO_2$) and oxygen consumption ($\dot{V}O_2$), see equation (A3.). Therefore, equation (A2.) could be rewritten in similar from as equation (A1.). However, again caution should be taken here, since the measured $RER$ signal, given by the indirect calorimeter software, slightly differ from the values that would result from manual computation of the $RER$ from the oxygen and carbon dioxide rate using equation (A3.), see Figure A1.

$$RER = \frac{\dot{V}CO_2}{\dot{V}O_2} \qquad (A3.)$$

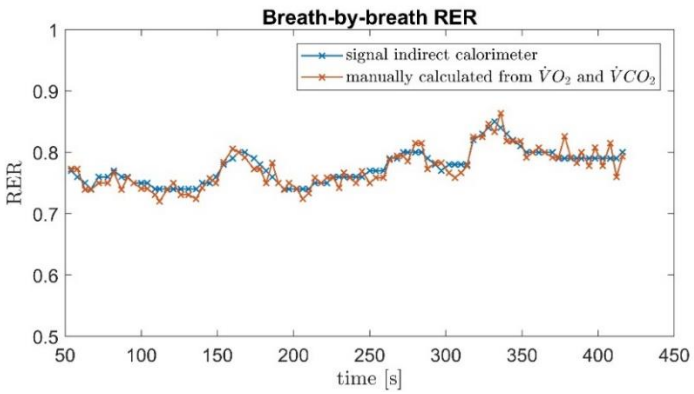



**Figure A1**. Illustration of difference in respiratory exchange ratio (RER) value between the signal from the indirect calorimeter and the manually calculated RER value from the oxygen and carbon dioxide rate, measured during a rest metabolism measurement.

It should be noted that the difference in units of equations (A1.) and (A2.) can be solved by dividing the output of equation (A2.) by 60, since the output $J \cdot min^{-1}$ than becomes similar as Watts ($J \cdot s^{-1}$). Nevertheless, there are still differences in ratio between the given constants in the equations, resulting in a small difference in the mEE output. This difference is caused by a difference in stoichiometry of the dietary species used by Brookway (1987) and Garby and Astrub (1987) during their glucose and lipid oxidation experiments. Based on the ratio of the saccharides and the length of the fatty acids in the chosen dietary species, a unique heat of combustion can be computed for carbohydrates and lipid oxidation (Kipp et al., 2018). According to Kipp and colleagues (2018) the equation is irrelevant within study design comparisons but do matters when comparing mEE values across studies. Nevertheless, both equations are only true for a steady-state metabolism.

#### 9.1.2. Subtraction of the rest metabolism

The outcome of the rest measurement is commonly used to subtract the rest metabolism from the *gross* energy expenditure during the walking test which result in the *net* mEE of gait (Brehm et al., 2007), see equation (A4.). The net mEE is assumed to give a more direct indication of gait efficiency since it becomes independent of the rest metabolism and in this way reduce the inter-participant variability (Brehm et al., 2007). However, the net mEE is more prone to variability due to variation or measurement errors in the rest mEE (Brehm et al., 2007; Schwartz, 2007).

$$mEE_{net} = mEE_{gross} - mEE_{rest} \qquad (A4.)$$

#### 9.1.3. Mass normalization

The simplest and most widely used normalization of mEE is the mass normalization (Schwartz, Bourke, Baker, 2005). Wherein the gross or net mEE are normalized for the body mass [kg] of the participant resulting in the gross or net metabolic energy *consumption* (mECS) [$J \cdot min^{-1} \cdot kg^{-1}$] or [$W \cdot kg^{-1}$], see equation (A5.). Mass normalization makes the outcome more valid for between-subject and long-term evaluations (Brehm et al., 2007; Swartz et al., 2007). However, again caution should be taken by measuring the body mass, since the body mass can differ between conditions, depending on the worn attributes (e.g. clothes, shoes, wearable walking device). The present study applied mass normalization to make the results comparable to the clinically determined steady-state values.

$$mECS = \frac{mEE}{body\ mass} \qquad (A5.)$$

#### 9.1.4. Speed normalization

Moreover, since mEE is highly dependent on the walking velocity (Dal, Erodogan, Resitoglu, Beydagi, 2010) it is common to normalize the mECS for walking speed [$m \cdot s^{-1}$] which result in the metabolic energy *cost* (mEC) [$J \cdot kg^{-1} \cdot m^{-1}$] also called the *cost of transport* (COT), see equation (A6.). According to Plasschaert, Matthews, and Forward (1999) the mEC is more reliable and sensitive for gait efficiency comparisons compared to the mECS. However, caution should be taken by the measurements of walking speed, especially when using a self-paced walking speed.

$$mEC = \frac{mEC}{Walking\ Speed} \qquad (A6.)$$

#### 9.1.5. Effect of the different methods on the subject-specific time constant

As mentioned in the introduction, the constant is a result of the model fit and the values depend on the data were the constant is fit on. This explains the difference between the smaller oxygen kinetics values $\tau_{\dot{V}O_2}$ = 30s reported by Potter & Unnithan (2005) compared to the $\tau$ = 42s found by Selinger and Donelan (2014). The present study explored these differences and found that a fit to the oxygen rate indeed result in lower time constants compared to a fit to the energy expenditure. Moreover, very

small differences were found between the time constant fitted to the energy rate data calculated using the equation of Gabry & Astrub (1987) and the equation of Brookway (1987). These differences in methodology should be taken into consideration when comparing among studies but also when deciding to use a recommended standard value from literature.

**Table 1A. Differences in subject-specific tau among different calculation methods**

| | | *Subject-specific tau* | |
|---|---|---|---|
| *Method* | | *M* | *SD* |
| Garby & Astrub (1987) | [J·kg$^{-1}$·min$^{-1}$] | 41.5 | 12.9 |
| Brookway (1987) | [W·kg$^{-1}$] | 39.9 | 12.8 |
| Oxygen rate | [L·kg$^{-1}$] | 41.4 | 13.0 |

### *9.2. Estimation-error-measurement-duration relationship among groups*

Because the sizes of the groups in the present study were rather small and unequally sampled, the present study focused in the results section on the total population. However, it might be interesting to visualize the results for the adult patients (i.e. Multiple Sclerosis and Neuromuscular Disease n = 23) and typically developing children ($n$ = 4) as well to compare with the results of the healthy adults from literature.

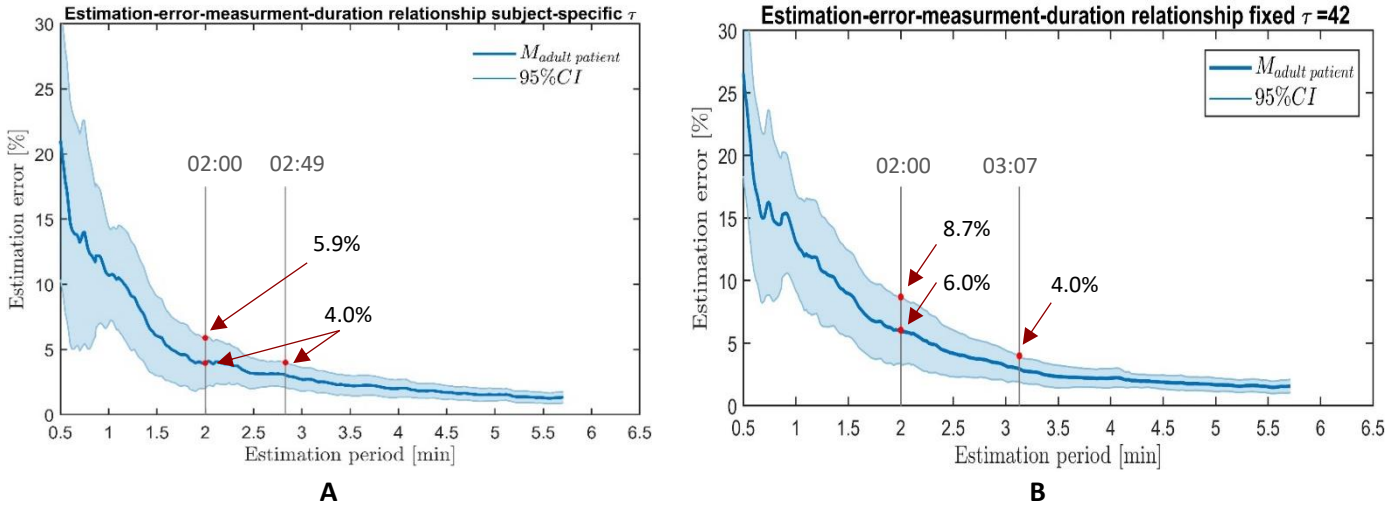



**Figure A2**. Estimation-error-measurement-duration relationship of A) the adult patients in the included population ($n$ = 23). The figure shows that after 2 minutes of measurement there is an average error of 4.0%, but a 5.9% error for the upper 95% confidence interval boundary. Before a 4% error level is reached in 95% of the subjects the measurement should take 2m49s. In B) the results for the general model in the adult patients of the population.

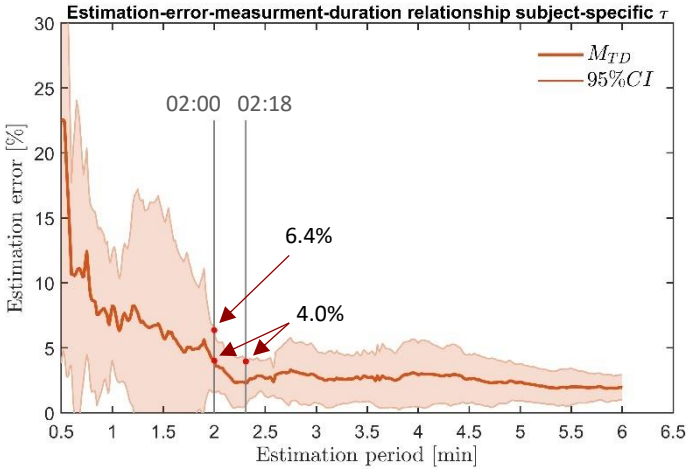



**Figure 25**. The estimation-error-measurement-duration relationship of the typically developing (TD) children in the included population ($n$ = 4). The figure shows that after 2 minutes of measurement there is an average error of 4.0%, but a 6.4% error for the upper 95% confidence interval boundary. Before a 4% error level is reached in 95% of the subjects the measurement should take 2m18s.

## *9.3. Strict inclusion criteria considerations*

The collected data contained a lot of trials where it was clearly visible that the participant did not reached an steady-state metabolism. However, since it is difficult to interpret raw breath-by-breath data, this study designed an objective method to make a selection of the most clean data for the explorative study. It might be the case that a few trails were excluded from analysis while they would have been 'clean' enough. However, after many considerations of different inclusion criteria the present study decided to use the currently used inclusion parameters, which extracts the most false positive appropriate for a first explorative study.

### 9.3.1. Steady-state determination criteria parameters

Considering the steady-state inclusion criteria, it might be the case that a moving window of 180 s recommended in the study of Schwartz (2007) is relatively large. The effect of different window lengths was inspected, it was found that a window of 120s included trials ($n$ = 67) was clearly not stable based on visual inspection of the data. A window of 160s ($n$ = 61) seemed more appropriate compared to 120s. With this window length a few trials were included that seemed appropriate based on visual inspection that were excluded by a window of 180s, however it still introduced an outlier in the final included dataset, see FigureA3A. Nevertheless, Schwartz (2007) investigated this window length and has well considered to use a window length of 180s, which resulted in the clearest included trials in the present study. Therefore, the present study decided to stay with this value.

The same holds for the rejection $\alpha$ level. Schwartz (2007) mention that an $\alpha$-level of .05 or .01, would accept more mildly rising or falling windows as stable. For examples of trials that would have been included using loosener inclusion parameters can be found in Figure A3B.

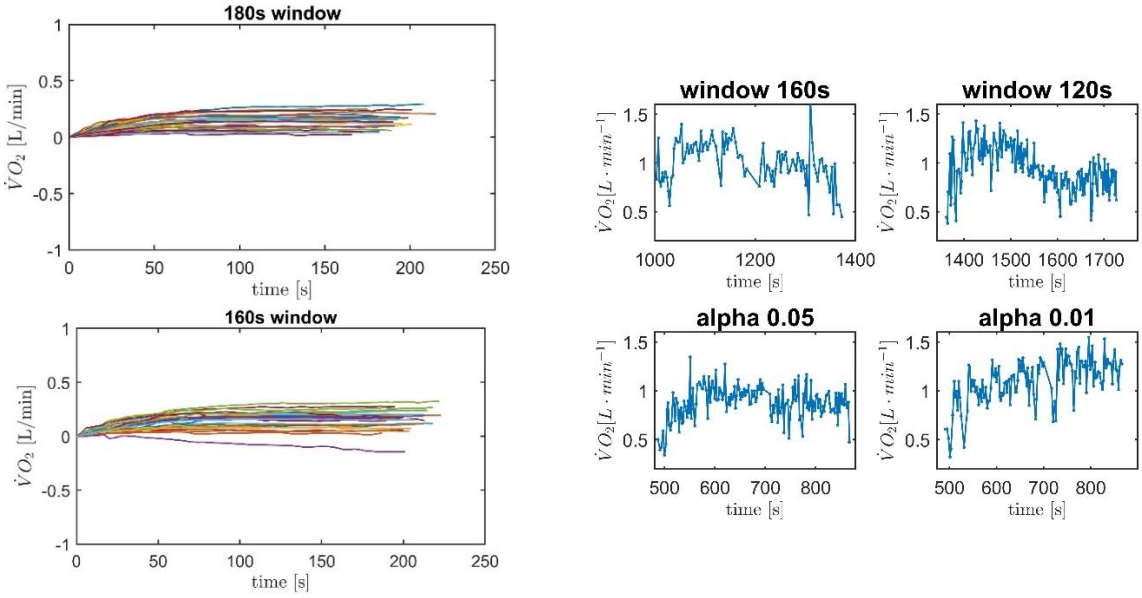



**Figure A3** . Examples of outliers in the included trials when using A) a window length of 160s compared to a window length of 180s.

9.3.2. Rest metabolism criteria

Concerning the rest metabolism inclusion criteria, it might be the case that it is not necessarily required to start from a rest metabolism to use the ICM model, however since the present study aimed to determine the subject-specific time constants it was required to assure that the metabolism starts from a steady value to another steady value (Selinger et al., 2014; Felt et al., 2015). Unfortunately, a gross amount of the collected trials did not start from rest. This is probably because it is common in the hospitals to measure the rest metabolism seated on a chair at a different location then where the over-ground track or treadmill is placed. Resulting in that subjects needs to walk after the rest measurement towards the track or treadmill location. In the meantime, the participant can move, talk and laugh in between the measurements, after which the participants start the walk test regardless of whether the patient's metabolism is fully returned to a rest value. This was probably more practical from a clinical perspective. In the conventional methods this would not affect their research outcome since they only use the steady-state metabolism at the end of the measurement. When the ICM technique would be used clinically in the future, these kind of small adaptations in the protocol designs should be considered.

9.3.3. Walking speed criteria

Concerning the walking speed criteria, this study initially aimed to meet the requirement of a step change in workload, after which the speed remains constant. This is aimed because the ICM model make use of this step change time response. After collecting and exploring the data it became clear that the data also contained trials that were not reaching a stable metabolism, see Figure 4A. However, since it cannot be seen from the mEE whether this is a response to a changing workload an attempt was made to inspect the walking speed consistency. But unfortunately, the lap-times were not available and/or not provided in 3 out of the 7 studies (i.e. measured in 2006, 2014 and 2016). Therefore, already 77 trials were excluded. Of the remaining trials an estimation could be made concerning the consistency of the walking speed. But it turned out that still in many trials walking speed was not consistent. At least in a much lower number of trials than expected (see Figure 5A for an example), raising questions concerning the correctness of the clinically outcomes of these trials.

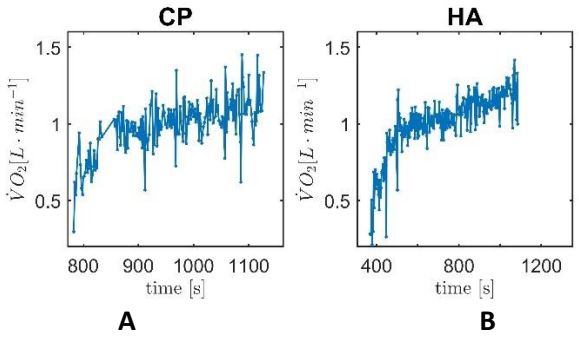



**Figure A4.** Two example of trials that are not reaching a steady metabolism of A) a child with Cerebral Palsy, and B) a healthy control subject.

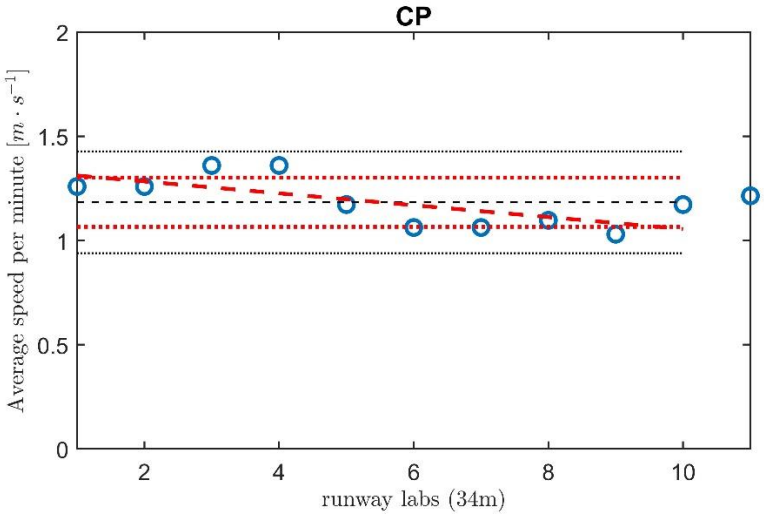



**Figure 5A**. The walking speed consistency over the different laps of the runway in a child with Cerebral Palsy. The red lines represent the rejection based on the regression slope and deviations of the mean. The black dotted lines represent the mean and the range of the lap speeds of the participant.

### *9.4. Lack of information of the provided data*

During data collection many unforeseen challenges had to be overcome (e.g. no back-up files of raw data available, hardware/software problems of the indirect calorimeters to re-extract data, missing event markers, missing walking speed information, etcetera). It turned out to be difficult for the hospitals to keep track of the previous collected data in a structured and uniform manner. It was cumbersome to trace done what protocols were used in the performed studies of the provided data. Details such as: what sequence in protocol and/or what track distances were used, whether patients walked on shoes or with AFO's and/or walked with additional helping aids, were not easily accessible.

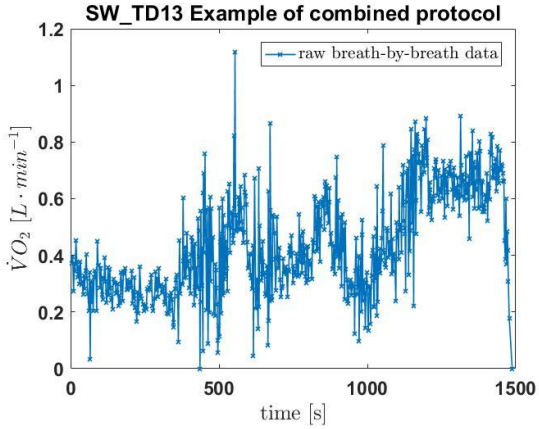



**Figure 6A**. Illustration of provided data where it in first instance was unclear what events took place during the protocol. Later, it turned out that the patients performed a calibration measurement in between the rest and walk measurement.

*9.4.1.1. Missing events markers*

In some trials, the information about the walking events, indicated with markers, were missing. It is quite challenging to interpret raw breath-by-breath results without any knowledge of the used protocol sequence and lack of information of the time point where upon the walking test must have started, see Figure 6A. These markers are normally synchronized with the metabolic data in the indirect calorimeter software and indicate the moment where upon the subject start the rest measurement, end the rest measurement, starts walking and stops walking. These events are crucial for the prediction since that determines when the external load on the body changes and when the prediction should start. When the used protocol was known and one or all markers were missing an attempt was made to select the time points where the rest and walking measurement began and ended using the implemented $ginput$ MATLAB function based on visual inspection. See Figure 7A for two examples where the marker events were missing and implemented based on visual inspection.

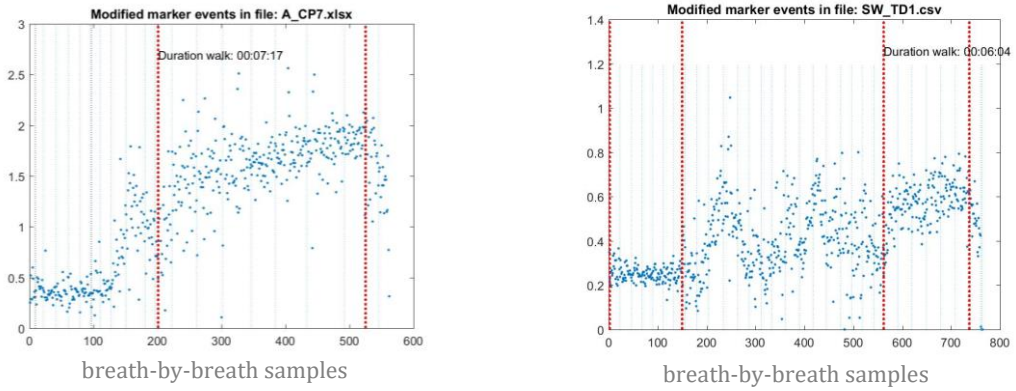



**Figure 7A.** Example of two trials were all markers were missing. With a visual support of vertical lines indicate the minutes in the data so that the duration of the rest and walking test could be counted. When a trial was too long it was automatically cut down to 6 minutes afterwards.

Especially in the older data (measured before 2016) this marker function was not often used because the software had no user-friendly button within the user interface at that time. No additional data log was provided with the data so retrospectively adding the true markers was not possible. This manually adding of the events markers based on visual inspection is not expected to have an effected on the results of the present study. In case of a too early placement of the marker the prediction might start too early, potentially slightly affecting the model fit and model parameter $\tau$, but since this takes

place in the first 10 seconds, this effect will be attenuate used measurement durations. In case of too late marker placements it could be the case that the trial is not included because the subject was not in rest anymore.

Moreover, it is found that the software of the indirect calorimeter places the marker, when the user clicks on the marker button, in the next following breath sample. Depending on this breath duration the marker will be a delayed. When this is a delay of only one breath sample this effect seems irrelevant. However, it might be the case that the user clicks too late on the button because it needs to perform a double task, such as give corrective instructions to the patient. Based on inspection of the original markers this seemed to be the case in some of the trials, see Figure A8. Unfortunately, the underlying cause is not completely known, it might also be due to neural anticipation of the coming exercise and therefore taken for granted. This might have resulted in exclusion of potentially includable trials.

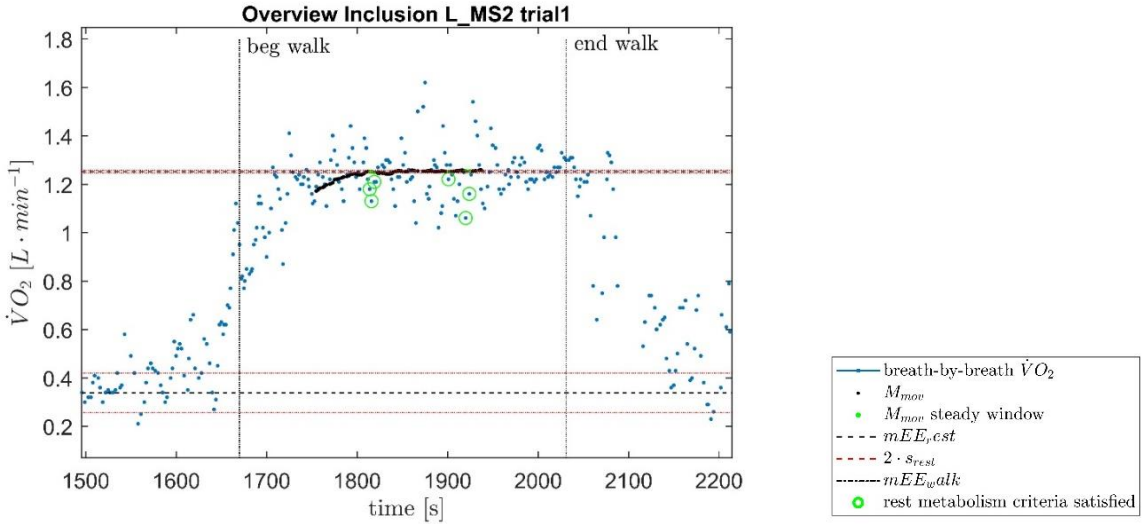



**Figure A8**. An example of a trial wherein the original placed marker seemed to be delayed compared to the true start walking event. Resulting in exclusion of the trial by the rest metabolism criteria.

9.4.2. Re-extracting raw back-up files of the data

The major challenge, and the most time-consuming part of the present study was the collection of the data. A correct and consistent maintenance of the data was lacking. The major issue started by the fact that the indirect calorimeter software applies additional filters on the data which make it cumbersome to back-up both raw as the processed data for the clinicians. This was both the case in the VUmc as AMC hospital. At the VUmc department all provided data was not available in a raw (breath-by-breath) version. Therefore, all trials needed to be re-extracted piecewise from the indirect colorimeter computers. This brought a lot of unforeseen difficulties. For example, the data was only available on two laptops with the old version of the software of the indirect calorimeter (METAMAX) that were used in the past to perform the gas analysis. These laptops were not used anymore and at first, it was unclear where these laptops were kept. After the laptops were found it seemed not possible to read the backup file that was made with the new software. The help of a software specialist of the indirect calorimeter was needed to figure this out. Moreover, the settings of the software were not correct causing difficulties to export the data in *.xlsx* files. Again, a software specialist was needed to reset these settings because no clinician in the hospital had the time or was capable in solving this problem. But since, there was no correct back-up made of the old datasets, it was not allowed to figure it out alone. Unfortunately, these hardware and software challenges took a lot of time reducing the time for the main analysis. At the AMC hospital a sort of similar problem was encountered, from the 393 provided trials only 171 trials were available in a raw version. Although the number of available trials was great, still a lot of trials could not be included. In this case, the raw version could not be re-extracted from the device since the indirect calorimeter laptop was no longer available.

*9.5. Treadmill belt speed GRAIL*

The present study has assumed that when a person starts walking from still standing on an over-ground track this could be considered as a step change in workload. In the MOVING-MS study, that used pre-defined walking speed performed on the GRAIL, it was evident that the speed was consistent over the walking trial. However, the treadmill has an internal algorithm to accelerate the belt speed towards the set target speed, allowing participants to habituate to the target speed, instead of applying an abrupt change.

This slower acceleration of the treadmill could have affected the increase in initial instantaneous metabolic energy demands of the body (possibly not meeting the required step change in workload). This holds for both the MOVING-MS as the MSWALK-I study. However, as far as known, this effect has not been mentioned in the found human-in-the-loop optimization literature that use also treadmill protocols. Probably because the internal acceleration of the treadmill is neglectable. However, the present study investigates trials that were measured in patient populations with walking impairments and possibly decreased stability. The experimental operator might have manually increased the treadmill belt speed with a slower acceleration. This might deviate from a step change potentially approaching more a *ramp change*. What would require a different type of time response of the IMC model. Therefore, the attempt was made to measure the acceleration of the treadmill belt in the study of MOVING-MS, that took place simultaneously with the present study, to inspect whether the acceleration of the treadmill could be considered as a step change. This difference could only be assessed by inspection of the average speed of the first minute compared to the first lap-speed, see Figure A9.

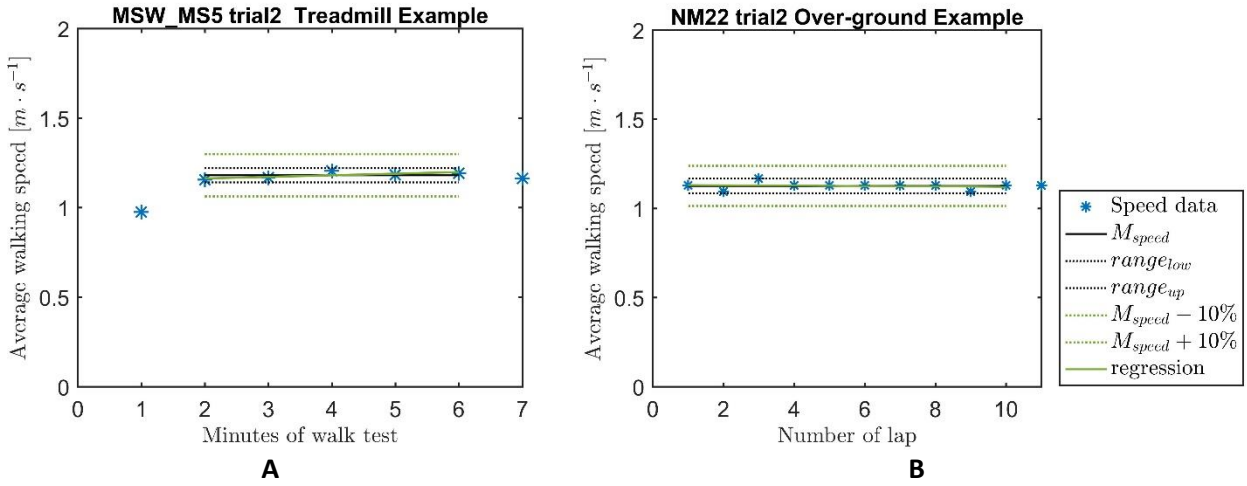



**Figure A9**. Illustration of two trials wherein the walking speed was consistent over the walking duration. The walking test of A) took place on a treadmill and B) is performed over an over-ground track. The average walking speed is plotted per A) minute of the walk test and per B) successfully completed laps. In A) the first walking speed is lower and outside the inclusion boundaries ($M \pm 10\%$) due to the lower acceleration of the treadmill. The first minute of the treadmill example was not included into the inclusion procedure (illustrated with the green area).

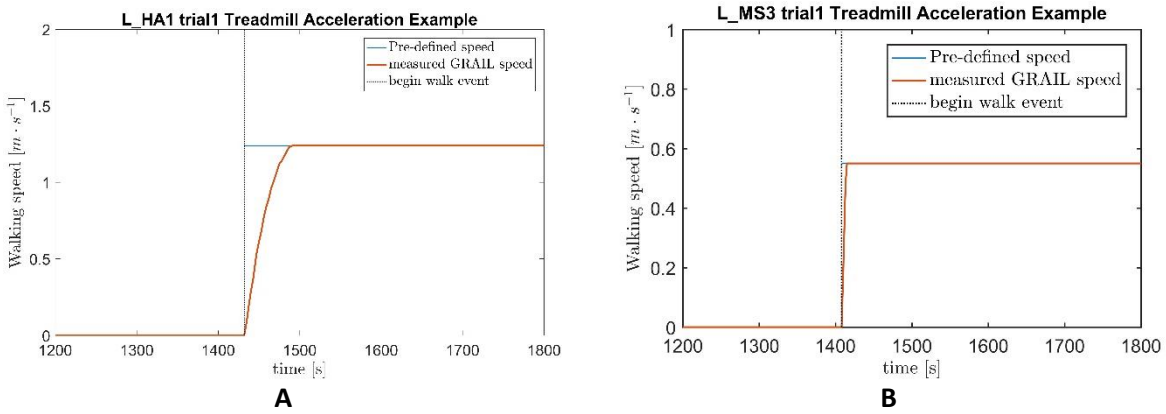



**Figure 26**. Two examples of the acceleration of the treadmill wherein in one trial A) the experiment operator manually accelerated the treadmill in a stepwise manner, and B) the internal acceleration of the treadmill is used. Whether the manual or internal acceleration was applied depended on the preference of the experiment operator.

After exploration of the acceleration of the treadmill that was available in a few trials (of the MOVING-MS study that was performed simultaneously with the present study), the present study choose to consider the change in workload within the treadmill studies as an approximated step change and did not take the walking speed of the first minute into consideration with the inclusion procedure. This might have resulted in larger determined subject-specific time constants for the MS and HA population groups. However, the present study did not identified differences in subject-specific time constants suggesting that this has no significant effect. However, the groups were small, and an effect of the treadmill might have been overseen.

Moreover, for the MS-WALK-I study an attempted was made to collect the raw treadmill belt speed over the whole test. However, the provided raw data seemed not consistent with the previous calculated average walking speed per minute that were calculated by the original researchers of the MSWALK-I study (MSc Thesis: Geurtz, & Los, 2016). Moreover, the treadmill belt speed was not available over the total duration of the test since the GRAIL system measures a lot of variables including 3-Dimensional kinematic data which result in enormous data files which are cumbersome to process. Therefore, separate intervals of approximately one minute are measured and these intervals are used to compute the averaged walking speed of approximately each minute within the walking test (MSc Thesis: Geurtz, & Los, 2016).

### *9.6. Comparison with clinical values*

In 10 participants the clinically determined values for the steady-state energy expenditure was available in the same form (no normalization by walking speed) and calculated according to the similar equation of Garby and Astrub (1987). An additional test has been performed to identify differences among the clinically determined values and the results of the present study. On average, the objectively determined $mEE_{walk}$ ($M = 321.52, SE = 31.20$) was slightly smaller than the clinically determined steady-state energy expenditure ($M = 322.30, SE = 30.49$). But the parametric one-sampled T-test revealed that this difference of .78 was not significant, $t(9)$ =.03, $p$=.98, $d$= .01. The ICM model outcome $\dot{\bar{E}}$ ($M$=325.47, $SE$=32.74) was slightly larger than the clinically determined steady-state energy expenditure. But again, this difference of -3.17 was not significant, $t(9)$ = -.10, $p$=.92, $d$=-.03.

### *9.7. Consistency of subject-specific time constant*

When multiple trials within a subject were available, the consistency of the subject-specific $\tau$ was evaluated. This was the case in a small group ($n = 7$) whereof two trials were included. The K-S test shows that the subject-specific time delay $\tau$ determined in the first trial $D(7) = .20, p > .05$, and

second trial $D(7) = .16, p > .05$ did not deviate significantly from normal. On average, the subject-specific time constant $\tau$ of the first trial ($M$ = 40.38, $SE$ = 4.65) does slightly differ from a second trial ($M$ = 33.04, $SE$ = 4.55), however, the Paired-Sampled T-test revealed that this difference of 7.33 is not significant, $t(6) = 1.45, p = .10, d = .55$.

### *9.8. Can smoothing improve the estimation-error-measurement-duration?*

The rehabilitation departments of the Amsterdam UMC both handle additional filters over the data prior processing. This is performed in the indirect calorimeter software. For the main analysis the raw breath-by-breath data was used to prevent a possible bias and circular reasoning. First the correctness of the ICM model itself needed to be explored. Nevertheless, it might be interesting to evaluate whether such additional processing of the data has an effect on the ICM model outcomes. Therefore, the similar data processing of the AMC hospital is applied over the data and the effect of the filter are evaluated for both the subject-specific modal as the general model. The AMC use a smoothing filter with a span window of 5s this was retrospectively applied to all included data by the implemented $smoothing$ function. It was expected that additional filtering would not have a great impact on the results since the ICM model itself has low pass filter characteristics. From the results it can be concluded that additional filtering has barely any effect on the estimation-error-measurement-duration relationship. However, it smoothens some alterations compared to the raw estimation.

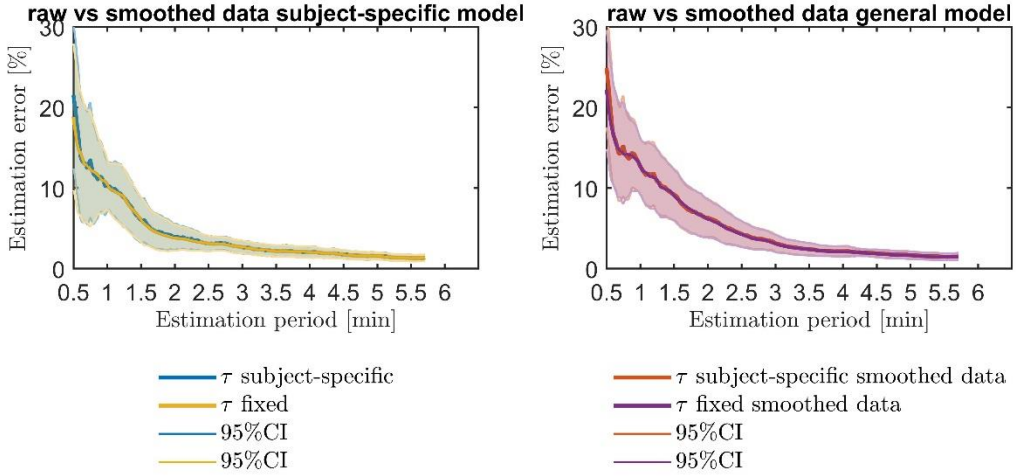



**Figure 27**. Comparison of raw and smoothed data on the subject-specific and general estimation-error-measurement-duration relationships.

It was found that using the full measurement duration, the subject-specific model had an individually correctness of 96% (i.e. 28 out of 27) using the smoothed data. This is similar to the results of the raw data. The general model showed a an individually correctness of 93% (i.e. 26 out of 27). This is one correct estimation more than with the raw data. These results suggest pre-processed data with additional smoothening will probably not harm the results of the ICM model and could potentially be used as well. Future research is needed to investigate this in more detail.