Lieferschein

Universitaetsbibliothek Regensburg

- Dokumentlieferung –
Postfach Postfach

D-93042 Regensburg

Tel.: ++49-941-943-3943
Fax: ++49-941-943-1958
Email: subito.UB@bibliothek.uni-regensburg.de

Empfänger
Helmholtz Zentrum Muenchen GmbH

Zentralbibliothek / Fernleihe

D-85764 Neuherberg

Ingolstaedter Landstr. 1

Angaben zur Bestellung:

Bestelldatum:	2017-02-08 13:47:34	
Bestellnummer:	SUBITO:VE17020800885	E000010655
Name des Bestellers:	Helmholtz Zentrum Muenchen GmbH	
Benutzerkennung:	SLS02X00668	
Lieferdatum:	2017-02-08 15:15:00	
Lieferpriorität:	NORMAL	
Aktueller Lieferweg:	Email	
E-Mail Adresse:	library@helmholtz-muenchen.de	

Bemerkungen zur Auslieferung:

Angaben zum Dokument:

Signatur:	[ONL] ND:2030773-1
Autor:	
Titel:	Methods of information in medicine
Jahr:	2016
Band / Jahrgang:	55/6
Seiten:	525-532
Aufsatzautor:	van, Hees
Aufsatztitel:	Challenges and Opportunities for Harmonizing Research Methodology
ISSN:	
ISBN:	
CODEN:	

Ihre Bemerkung zur Bestellung: Paulini

subito Urheberrechtshinweis

Die Bestellung und Nutzung der über subito gelieferten Aufsatzkopien unterliegen den urheberrechtlichen Bestimmungen. Mit der Registrierung bei subito verpflichten Sie sich, diese einzuhalten, d.h. insbesondere, dass die Kopien ausschließlich zum eigenen Gebrauch bestimmt sind und nicht an Dritte weitergegeben werden dürfen. Sie dürfen ohne Genehmigung des Verlags nicht zum Wiederverkauf, Wiederabdruck, zu systematischer Verteilung, Emailversand, Webhosting eingeschlossen institutionelle Repositorien/Archive oder jedweden anderen kommerziellen Zweck verwendet werden.

Sofern Sie eine Lieferung per Email oder FTP erhalten, dürfen Sie die Kopie nur einmal ausdrucken und müssen diese anschließend dauerhaft löschen.

Die Kopien sind mit einem Wasserzeichen versehen, welches ein Urheberrechtsvermerk enthält. Das von subito e.V. angebrachte Wasserzeichen darf nicht entfernt werden.

Challenges and Opportunities for Harmonizing Research Methodology: Raw Accelerometry

Vincent T. van Hees[1]; Kathrin Thaler-Kall[2]; Klaus-Hendrik Wolf[3]; Jan C. Brønd[4]; Alberto Bonomi[5]; Mareike Schulze[3]; Matthäus Vigl[6]; Bente Morseth[7,8]; Laila Arnesdatter Hopstock[7]; Lukas Gorzelniak[5]; Holger Schulz[9]; Søren Brage[10]; Alexander Horsch[2,11,12]

[1]Netherlands eScience Center, Amsterdam, The Netherlands;
[2]Institute of Medical Statistics and Epidemiology, Technical University of Munich, Klinikum rechts der Isar, Munich, Germany;
[3]Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig – Institute of Technology and Hanover Medical School, Hanover, Germany;
[4]University of Southern Denmark, Odense, Denmark;
[5]Philips Research Laboratories, Eindhoven, The Netherlands;
[6]German Institute of Human Nutrition Potsdam-Rehbrücke (DIfE), Potsdam, Germany;
[7]Department of Community Medicine, UiT – The Arctic University of Norway, Tromsø, Norway;
[8]School of Sport Sciences, UiT – The Arctic University of Norway, Tromsø, Norway;
[9]Institute of Epidemiology I, Helmholtz Zentrum München, Munich, Germany;
[10]Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom;
[11]Department of Computer Science, UiT – The Arctic University of Norway, Tromsø, Norway;
[12]Institute of Clinical Medicine, UiT – The Arctic University of Norway, Tromsø, Norway

Keywords
Standardization, accelerometer, physical activity, raw data

Summary
Objectives: Raw accelerometry is increasingly being used in physical activity research, but diversity in sensor design, attachment and signal processing challenges the comparability of research results. Therefore, efforts are needed to harmonize the methodology. In this article we reflect on how increased methodological harmonization may be achieved.
Methods: The authors of this work convened for a two-day workshop (March 2014) themed on methodological harmonization of raw accelerometry. The discussions at the workshop were used as a basis for this review.
Results: Key stakeholders were identified as manufacturers, method developers, method users (application), publishers, and funders. To facilitate methodological harmonization in raw accelerometry the following action points were proposed: i) Manufacturers are encouraged to provide a detailed specification of their sensors, ii) Each fundamental step of algorithms for processing raw accelerometer data should be documented, and ideally also motivated, to facilitate interpretation and discussion, iii) Algorithm developers and method users should be open about uncertainties in the description of data and the uncertainty of the inference itself, iv) All new algorithms which are pitched as "ready for implementation" should be shared with the community to facilitate replication and ongoing evaluation by independent groups, and v) A dynamic interaction between method stakeholders should be encouraged to facilitate a well-informed harmonization process.
Conclusions: The workshop led to the identification of a number of opportunities for harmonizing methodological practice. The discussion as well as the practical checklists proposed in this review should provide guidance for stakeholders on how to contribute to increased harmonization.

Correspondence to:
Prof. Dr. Alexander Horsch
Department of Computer Science, UiT – The Arctic University of Norway
Tromsø
Norway
E-mail: alexander.horsch@uit.no

Methods Inf Med 2016; 55: 525–532
http://dx.doi.org/10.3414/ME15-05-0013
received: December 6, 2015
accepted: June 7, 2016
epub ahead of print: October 7, 2016

Funding
This work was supported by the German Research Foundation (DFG).

1. Introduction

Accelerometry is increasingly being used in physical activity research and a recent review [1] highlighted the difficulties of comparing estimates from different accelerometer brands when these incorporate substantial on-board processing before data are stored on the device, advocating for increased use of raw sensor data, in line with our earlier suggestions [2]. Raw accelerometry refers to the sensor technology which stores acceleration data in units of gravitational acceleration or m/s^2 at sample frequencies of typically above 10 Hertz. A sketch of the data collection and processing pipeline that applies to raw data accelerometry is shown in ►Box 1. The technology

Schedule

The Raw Data Accelerometry Collection and Processing Pipeline

1. Research question, including a specification of what aspects of physical activity are of interest
2. Use literature, new empirically study, and/or expert advice to identify the most appropriate accelerometer, study protocol, and accelerometer configuration to capture these aspect identified under #1
3. Study itself, the study participants wear sensors, data capture and storage
4. **Raw data**
5. Extraction of signal features needed for steps 6, 7 and 8. For example filter the signal and take the average per second.
6. Use signal features to inspect and improve data quality, e.g. calibration to gravity and protocol compliance by participant. Step #6 is used to improve steps #7 and #8.
7. Describe the signal features and/or use signal features for inference algorithms to classify or estimate something that was not directly measured like activity types or energy expenditure
8. Physical activity outcome variables, typically the weekly or daily average or total of the information extracted under #7

of raw accelerometry has been around since the 1990s but has only become feasible for the use in the assessment of weeklong human physical activity since the first decade of this century [2, 3]. The introduction of this technology to the field has fueled the development of methods to manage, analyze and interpret raw accelerometer data [4–6]. The access to raw sensor data, and the consequential analytical freedom facilitates methodological innovation, but may paradoxically also challenge methodological consistency [1, 7]. Another challenge for methodological consistency is that raw accelerometer data may not be 100% compatible between accelerometer brands [8, 9]. Methodological consistency is critical for valid comparisons of physical activity data from different populations, from clinical to nationally representative population samples, and at different points in time [2, 10–12]. Although method harmonization could in theory be as simple as selecting the most popular method from the scientific literature and advising everyone to use that method, this is an unrealistic proposition because:
i) there is already a great diversity of methods in use, ii) practical requirements differ between method users, by which one standard may not suit all, and iii) future deviations from an established set of standards may represent innovation which should be encouraged. Therefore, efforts are needed to harmonize various aspects of methodological practice which should entail both the consideration of consistency as well as that of methodological innovation.

The GSISH biosensor workshop is a recurring workshop initiated and organized by the Technical University of Munich (TUM) Graduate School for Information Science in Health (GSISH) [13] with invitees from the field of physical activity assessment. In March 2014, the third edition of this workshop took place and was themed around the possibility of establishing standards for raw accelerometer data methods. The workshop attendees are the authors of this article, representing expertise in sensor manufacturing [14], signal processing [15], study design [16], doubly labelled water studies [4, 17], clinical studies [18, 19], and population-based research [20–25].

The aim of this article is to use the discussions from the workshop to reflect on how methodological harmonization can be achieved and to compile practical check lists to guide harmonization.

2. Methods

The workshop included plenary discussions as well as subgroup discussions. First, generic aspects of the harmonization process were discussed: (i) Who are the stakeholders? (ii) How are methodologically standards typically established? and (iii) How does the harmonization topic relate to the fundamental principles of scientific transparency and openness? Next, each aspect of raw accelerometry methodology was discussed: i) Hardware, ii) Data collection and study protocols, iii) Method development and evaluation, iv) Data processing, v) Data description before and after processing, vi) Method implementation and evaluation.

For each of these components we consider opportunities and challenges with respect to how harmonization can be established and what roles and responsibilities various stakeholders can take on.

3. Results

3.1 Generic Aspects about Harmonization

3.1.1 Stakeholders

Harmonization relies directly on three key stakeholders: sensor manufacturers, algorithm developers, method end-users (see ►Figure 1). It should be noted that individuals can be part of more than one of these three groups; there are sensor manufacturers who develop and release algorithms and there are end-users who also contribute to algorithm development. Method end-users are those who apply developed methods in order to then study the association of the information that was derived from the raw acceleration data with some other quantity, for example the occurrence of a disease [26] or underlying population demographics [22]. In addition to these three groups of stakeholders, there are other parties involved, including funders, scientific journals, toolkit/advisory websites, and public platforms for data and algorithm sharing (see ►Figure 1), who play more indirect roles in this process by supporting and/or endorsing methodological practices.

3.1.2 Evolution of Methodological Standards

A methodological standard should ideally evolve naturally by ongoing accumulation of empirical evidence on the feasibility and validity of a method. However, standards often evolve as a result of being the first of their kind or because large studies opt to use particular methods, e.g. the use of count-based uniaxial waist-worn accelerometry in 15,000 individuals in the NHANES 2002–03 and 2005–06 cycles [11, 27], or the choice of raw triaxial wrist-worn accelerometry in over 100,000 individuals participating in UK Biobank [10]; this may have influenced other studies to also opt for very similar methods [23, 28]. Regardless of the justification for choice of any method, one cannot deny that it is highly attractive to copy the study design of such large studies, as it guarantees comparability and facilitates the availability and the exchange of analytical methods. The risk with this approach to adopting a methodology is that the particular method may become so popular that its continued evaluation may be discouraged. However, it is widely acknowledged that the validity of a methodological approach is not only tied to the raw data as collected, and to the study protocols, but very much also to the inference algorithms that are used to process the data and the population to which the inference is applied. This leaves two important goals of continued evaluation of existing methods: one is development of new algorithms for the same raw data, and the other is the evaluation of validity of proposed algorithms in different populations or settings. On the other hand, the willingness of large-scale studies to implement novel technologies can be an important driver of methodological innovation as it puts the limitations of a technology right in the spotlight [8, 29–32] and by that encourages efforts to address those limitations.

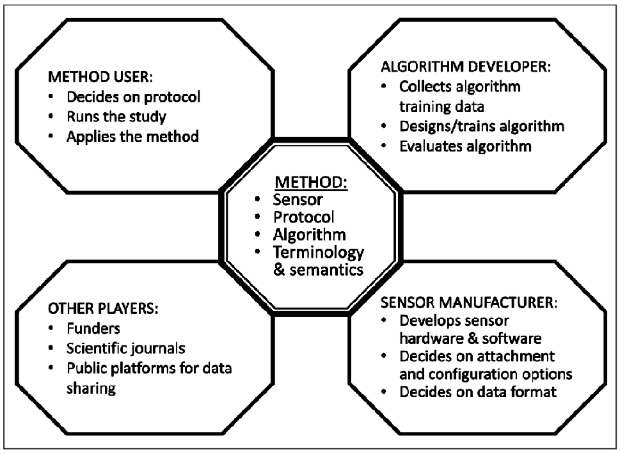

Figure 1 Diagram to visualize the main stakeholders that can contribute to the harmonization of methods for raw data accelerometry

3.1.3 Transparency

It was proposed that accessibility and transparency should be the guiding principle for methodological harmonization. However, complete transparency may not always be possible or realistic to achieve. Commercial parties may not disclose all information related to their products in order to give them a commercial advantage and protect their investments to ensure profits; the challenge for companies is to provide enough information to their end-users about the nature of the data that is being generated, coupled with timely identification and filing of patents required to prevent other companies from simply copying their products. It should be noted that for the end-user a not fully transparent method is not necessarily a worse method, although some reassurance of the method's applicability to a given study's primary aim is warranted. In research settings where resources are limited, it may in some instances be more cost-effective to rely on commercial services for derivation of required variables than to go through the potentially high burden of doing all the interpretative work in-house in an open fashion.

3.2 Specific Components of a Method

3.2.1 Hardware

Accelerometers from a variety of manufacturers have been used to collect raw accelerometer data, including among others: GENEA, Unilever Discover, Ltd, UK [2], GENEActiv, ActivInsights Ltd, UK [23], Axivity, Axivity Ltd, UK [33], and Actigraph, Actigraph LLC, USA [9]. These accelerometers would include a sensory element that performs the actual measurement of acceleration, and this sensor element may come from different manufacturers or come in different versions with different measurement properties. Comparisons between different accelerometer brands in terms of what is actually being measured and how are therefore not always straightforward. Manufacturers of accelerometers can contribute to increased transparency about their monitoring products by publishing sensor specifications. A suggested minimum level of information about sensor specifications is presented in ▶ Table 1. However, manufacturers may argue that aspects other than those related to methodological consistency drive the choices made during the product development process; combining the need for higher accuracy, and also physical requirements such as small size, long battery life, and flexibility of operational settings still represent a challenge that often leads to singular hardware and software architecture. It would perhaps be simpler for manufacturers to provide warranty on well-established performance specifications, which is ultimately the responsibility of the scientific community to define. For example, such a performance specification may include a particular relationship (e.g. linear) between sensor output and mechanically generated standardized acceleration signals, or that there should be no or only a well-defined marginal difference between the output of the new and the preceding sensor version when tested under such controlled conditions of a mechanical movement. If there is a difference then evidence needs to be provided that the newest version of the accelerometer is more valid, and ideally some guidance on how to map old to new data and vice versa [2]. For raw triaxial accelerometry, there are abundant opportunities to check sensor response to static accelerations between −1 and 1g using

Table 1 Suggested minimum information an accelerometer manufacturer should provide about their sensor

	Information	Example
1	Sensor type	Tri-axial MEMS accelerometer
2	Acceleration sensor manufacturer	Analog Devices
3	Acceleration sensor model	ADXL345
4	Possible sampling frequencies	50, 75, 100 Hertz
5	Sampling mechanism	Based on internal clock in acceleration sensor
6	Data resolution	12 bit over a 16g dynamic range
7	Dynamic range of sensor	+/– 8g
8	Specification of on-board signal processing	Analogue low-pass filter at 30 Hertz with –3dB
9	Report of a test to demonstrate the relationship of the output of the accelerometer with reference acceleration signals	Experiment with a robot
10	Report of a test to demonstrate that the output of the accelerometer is comparable, or more precise, in capturing standardised movement compared with its preceding versions	Experiment with mechanical shaker device or a robot
11	Temperature stability of the response	Statics experiment (not moving accelerometer) at different temperatures

the Earth's gravitational field even when sensors have been deployed in the field [34, 35]. However, documentation of dynamic acceleration response beyond that range remains the premise of mechanical shaker experiments requiring specialist equipment, thus likely best placed with the manufacturer. In return, the scientific community should try to inform manufacturers on the required resolution in the acceleration amplitude and the desired accelerometer frequency response during the activity types that are of interest.

3.2.2 Accelerometer Configuration and Study Protocol

Once an accelerometer is purchased, the end-user needs to decide on measurement configuration (e.g. sample frequency, active channels) and study protocol (e.g. anatomical attachment location, measurement duration, and instruction to the participant). Raw data accelerometry has been implemented with sample frequencies as low as 10 Hertz [36] up to 100 Hertz for long-term monitoring [37]. Further, most studies collect data over seven days, e.g. NHANES [38, 39], the German National Cohort [40], and UK Biobank [10], while a nine day protocol has also been implemented to ensure that at least seven full days defined from waking up in the morning till waking up the next morning are recorded [41].

Decisions about accelerometer configuration are generally constrained by limitations of the technology. Here, one could follow the principle of collecting as much information as possible (e.g. highest sample frequency as possible), collecting at least as much information as in similar studies, or justifying the accelerometer configuration based on the required information content of the signal (e.g. Shannon-Nyquist theorem for selecting sample frequency [42]). Harmonization in accelerometer configuration may be achieved by providing a motivation for accelerometer configuration in publications, which may then contribute to a scientific discussion about the strength of the motivation. Subsequently, less appropriate configurations should become less popular if a convincing motivation for them is missing.

Decisions on the study protocol are often driven by the cost of the method, primary outcome measures, the method's ability to deliver those outcomes, feasibility in the population of interest, assumptions about behavioral variability (number of days of monitoring required for a certain level of reliability), and evidence or suggestions from the literature or colleagues about what protocol works best. As a result, the harmonization could indirectly be facilitated by ongoing efforts to reduce method costs, increase measurement capacity, improve method feasibility, and evaluation of the reliability of outcome measures in various populations [43–45]. Further, harmonization may be encouraged by setting minimum standards for the quality of operational procedures related to accelerometer configuration and study protocol. In ►Box 2 we propose a minimum checklist for operational procedures when working with accelerometers.

Checkpoints

Suggested Checklist for Operational Procedures when Working with an Accelerometer

1. Is there a record of: body site of attachment, accelerometer attachment procedure, and accelerometer wear instructions (including e.g. night time and water-based activities)?
2. Is there a record of relevant participant characteristics, e.g. handedness or leg length?
3. Does the amount of data collected match the intended measurement duration?
4. Is the accelerometer housing intact?
5. Are the accelerometer and attachment straps cleaned after every measurement?
6. Is the battery sufficiently charged?
7. Is the accelerometer correctly configured?
8. Is there a study-level record of the dates on which the accelerometer was supposed to be worn?

3.2.3 Choice of Data Processing and Inference Methods

Procedures for data management and quality control are often not reported in publications, but likely differ between research groups. Reported inference methods also vary widely, e.g. by their choice of summary metric [22, 46].

It was also acknowledged that data management and data quality control need to be considered and reported in detail, e.g. detection of sensor malfunction, check for acceleration sensor calibration [34, 35], and accelerometer non-wear detection. As a starting point for harmonization we have compiled a minimum checklist for data management and data quality control, see ►Box 3. Furthermore, data processing tools that are able to provide comparable output despite variations in accelerometer configuration should be encouraged; for example, average rectified acceleration over 5-sec windows would be similar for signals sampled at 20 and 100 Hz.

Similarly, inference methods for energy expenditure and activity type are also subject to ongoing debate. For example, some publications on inference methods give the impression of being in favor of machine learning approaches [11, 47], while other publications seem to prefer a more heuristic approach, like threshold methods, prioritizing an understanding of the method and its output over data driven classification performance alone [9, 16, 48, 49]. It is likely that several of these methods may enhance our understanding of human behavior, but from different perspectives; however, for this potential to be realized, we need to establish some ground rules that would facilitate intrinsic understanding of each method on its own, as well as comparisons between methods.

To increase harmonization in data processing methods, the following actions are potentially beneficial: i) An attempt should be made to motivate every computational step of an algorithm and the order in which they appear. ii) Algorithms that are pitched as "ready for application" need to be published, which means the full algorithm including optimized parameters and coefficients, not the programming code that was used to train the algorithm. iii) Method users should report on algorithm settings and parameters which are essential to understand generated results. A proposed checklist for reporting on data processing is provided in ►Box 4.

Checkpoints

Suggested Checklist for Data Management and Quality Control

1. Screen data for accelerometer non-wear periods
2. Screen data for implausible data points
3. Inspect data for acceleration sensor calibration error, e.g. relative to gravitational acceleration
4. Verify sensor attachment orientation if the algorithm depends on it
5. Keep a record of data cleaning at study level, e.g. individuals excluded
6. Keep a record of data cleaning at individual level, e.g. measurement periods included/excluded
7. Keep a record of the algorithm version(s) used

Checkpoints

Suggested Checklist for Reporting on Data Processing

1. Report source and developer of employed algorithm(s)
2. Report parameters and coefficient values used for the algorithm(s)
3. Report on programming environment specifications
4. Provide a written description and motivation of the key steps in the algorithm
5. Provide, where possible, a reference to other publications using the same algorithm
6. Provide, where possible, literature references for studies supporting the appropriateness of the algorithm for application under the conditions for which it is used

3.2.4 Data Description before and after Processing

Data description in literature is currently not consistent, in particular in the level of detail that is used to describe different types of data, e.g. activity type "sitting" compared with activity type "sitting and instructed not to move". Another area of possible harmonization is the description of data and semantics of data. Descriptions of data can exist in various forms and with different purposes. Here, we will focus on all data related to training, developing, evaluating, and implementing algorithms for activity type classification and energy expenditure estimation. A distinction is necessary between three levels of data description: i) Description of the data before it is used to train the algorithm, ii) Description of the data when it is used for algorithm training, and iii) Description of the real study data the algorithm is applied to. The description of the data before it is used for algorithm training should ideally be interpretable by an independent person at any point of time; this is mandatory to ensure a broad and appropriate use for future users. The description (labelling) of data at the stage of algorithm training, e.g. the activity types to be classified, may require a simplified version of the previously mentioned easy interpretable description and is usually down to the objectives and design of the algorithm. However, once the algorithm is applied in unconstrained free-living individuals' data, the algorithm should ideally be explicit about the description of its output. For example, we may classify an activity "walking", but we know that the classifier was specifically trained on an able-bodied adult person walking on a level hard surface at normal speed in the laboratory, wearing shoes, therefore specifically not in walking barefeet on sand, walking at very slow speed, under free-living conditions, walking while wearing heels, or walking by a young or old or disabled person. Therefore, the description (labelling) of algorithm output should reflect the consequential vagueness in the derived activity types (or energy expenditure construct) and the uncertainty of the classification itself (e.g. probability for sitting may be highest, but it could also be standing). Despite its scientific importance, such a detailed description may not always be practical. Therefore, method developers and users will have to find a balance between sufficient detail in their data description and meeting requirements of practicality, a balance which may be application specific.

A naming convention can help to provide structure to the data and to ease data merging between users, but should not be used as an excuse to omit other types of data description. For example, if pictures are taken of the experimental environment then they should be kept with the data such that future users of the data have the opportunity to gain a better understanding of how and where the data was collected.

In conclusion, it seems that harmonization in the description of data is not necessarily about standardizing the data description, but more about the completeness of information and explicit expression of known unknowns in the description, such that appropriate comparisons can be made between datasets and algorithms.

3.2.5 Method Evaluation and Ongoing Development

Various aspects of a method can be developed or evaluated with robot experiments, exercise laboratory experiments, and free-living experiments. Each of these study designs has its specific strengths and weaknesses, and it is advantageous to combine several study designs to study multiple aspects of how a method may perform. Collaboration between research groups with complementary expertise should make such multifaceted evaluations possible. ▶ Table 2 provides an overview of the three categories of study designs and their potential functions. Here, a lack of harmonization may even be desirable because it ensures that methods are exposed to a diversity of experiments. Ongoing efforts to identify new experimental designs for developing a method could result in better methods, and ongoing efforts to identify new experimental designs for evaluating a method could result in the identification of yet unobserved limitations of a method. Needless to say that method comparisons should only be done under standardized conditions where methods are concurrently applied within a single study population. The reporting of development and evaluation experiments could be harmonized to some degree as we have described above. Algorithm developers are encouraged to carefully consider the replication of each other's techniques for algorithm performance assessment to ease the comparison of results.

Finally, harmonization and the general effectiveness of method development can be improved by facilitating the interaction between method developers and method users, e.g. by means of workshops, symposia, conferences and online discussion forums. Further, funding bodies and scientific journals may allocate resources and space for method development and evaluation, including pilot-testing of new methods for feasibility in clinical and population studies. To foster the transition from established methods to new approaches, we encourage reporting of results from both new and established methods, where possible. This allows for comparisons of temporal and spatial differences in population levels of physical activity.

Table 2 Study designs with which an accelerometer method can be developed and/or evaluated

Characteristic	Robot experiments	Laboratory	Free-living
Conditions	Specific kinematic conditions	Standardized activity types or freely chosen activities within a constrained environment	Real life environment, with optional prescription of activities
Variability testing	Intra- and inter-device variability	Inter-device variability, (intra-device variability more difficult)	Inter-device variability
Observational reference	Video observation	Video observation	Questionnaire and logbook (video but less feasible)
Non-observational reference methods	Robot kinematics	Indirect calorimetry, force platform, or other wearable sensor system	Doubly labelled water method or other wearable sensor system
Technical evaluation of accelerometer	Noise analysis, on-board filter characteristics, can the device keep time, testing under extreme conditions	Can the device keep time, on-board filter characteristics, testing under extreme conditions	Can the device keep time, testing under extreme conditions

4. Discussion

Methodological harmonization in raw accelerometry is desirable from several perspectives, but should ideally be the result of a scientific debate and ongoing collection of empirical evidence of the validity of methods. The discussion and practical checklists in this review aim to provide guidance to the process of harmonization. In summary, we identified the following key action points: i) Manufacturers are encouraged to provide a detailed specification of their sensors. ii) Each fundamental step of an algorithm for processing raw accelerometer data should be motivated to facilitate interpretation and discussion. iii) Algorithm developers and method users should be open about uncertainties in the description of data and the uncertainty of the inference itself. iv) All new algorithms which are pitched as "ready for implementation" should be shared with the community to facilitate replication and ongoing evaluation by independent groups. v) We recommend a dynamic and continuing interaction between method stakeholders to facilitate a well-informed harmonization process.

Despite efforts to harmonize methodology, it is likely that a certain amount of methodological inconsistency will always prevail in physical activity research as a result of ongoing innovations and need for comparisons with historical data. In the past, physical activity researchers have tried to develop conversion equations to compare published reports on physical activity outcomes based on inconsistent methodology [50, 51]. A major limitation of such retrospective conversion attempts is that they add another layer of uncertainty to the interpretation of data because of failing to capture the complex underlying processes needed to guarantee perfect compatibility [2, 8]. Therefore, the re-analyses of the original data should be encouraged as much as possible [1]. In addition, research will be needed to understand the impact of methodological differences on typical physical activity outcome variables and how this impact can be accounted for. Study designs to facilitate a better understanding of the impact of methodological inconsistencies may include: convergent validity studies to compare entire methods, data re-analyses to compare algorithms, and repeated measurements to compare study protocols [52]. Finally, the practical guidelines as proposed in this article should help discourage unnecessary methodological inconsistencies.

In comparison to other narrative review articles published on physical activity methods in recent years, this article is the first to focus exclusively on the harmonization, considering the full information pipeline ranging from hardware to physical activity outcome variables, and to use a multi-stakeholder panel discussion aimed at establishing practical guidelines [1, 12, 27, 53, 54]. This article does not aim to identify the latest trends and issues in methodology, but much more aims to make timeless recommendations on how to achieve harmonization in scientific methodology.

In conclusion, the workshop led to the identification of a number of opportunities for harmonizing methodological practice. The discussion as well as the practical checklists proposed in this review should provide guidance for stakeholders on how to contribute to increased harmonization and to stimulate further debate and consensus finding in the interdisciplinary physical activity research community.

Author Contribution

All authors participated in the workshop that led to this manuscript. KTK took minutes at the workshop. VTvH drafted the manuscript. All authors commented on various versions of the manuscript. All authors read and approved the final manuscript.

Conflict of Interest

No competing interests are declared.

References

1. Wijndaele K, Westgate K, Stephens SK, Blair SN, Bull FC, Chastin SFM, et al. Utilization and Harmonization of Adult Accelerometry Data: Review and Expert Consensus. Med Sci Sports Exerc. 2015; 47(10): 2129–2139.
2. van Hees VT, Pias M, Taherian S, Ekelund U, Brage S. A method to compare new and traditional accelerometry data in physical activity monitoring. 2010 IEEE Int Symp "A World Wireless, Mob Multimed Networks." IEEE; 2010; 1–6.
3. Esliger DW, Rowlands A V, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. Med Sci Sports Exerc. 2011; 43(6): 1085–1093.
4. van Hees VT, Renström F, Wright A, Gradmark A, Catt M, Chen KY, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. PLoS One. 2011; 6(7): e22922.
5. Zijlstra W, Hof AL. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait Posture. 2003; 18(2): 1–10.
6. Veltink PH, Bussmann HB, de Vries W, Martens WL, Van Lummel RC. Detection of static and dynamic activities using uniaxial accelerometers. IEEE Trans Rehabil Eng. 1996; 4(4): 375–385.
7. van Hees V. The challenge of assessing physical activity in populations. Lancet. 2012; 380(9853): 1555; author reply 1555–1556.
8. Brønd JC, Arvidsson D. Sampling frequency affects the processing of Actigraph raw acceleration data to activity counts. J Appl Physiol. 2016; 120(3): 362–369.
9. Hildebrand M, Van Hees VT, Hansen BH, Ekelund U. Age-Group Comparability of Raw Accelerometer Output from Wrist- and Hip-Worn Monitors. Med Sci Sports Exerc. 2014; 46(9): 1816–24.
10. UK Biobank. Category 2 enhanced phenotyping at baseline assessment visit in last 100–150,000 participants [Internet]. 2009. p. 33. Available from: http://www.ukbiobank.ac.uk/wp-content/uploads/2011/06/Protocol_addendum_2.pdf.
11. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med. 2014; 48(13): 1019–1023.
12. Pedišić Ž, Bauman A. Accelerometer-based measures in physical activity surveillance: current practices and issues. Br J Sports Med. 2015; 49(4): 219–223.
13. GSISH website [Internet]. 2014 [cited 2014 Jul 23]. Available from: http://gsish.tum.edu/.
14. Bonomi AG, Goris AH, Yin B, Westerterp KR. Detection of type, duration, and intensity of physical activity using an accelerometer. Med Sci Sport Exerc. 2009; 41(9): 1770–1777.
15. Gyllensten IC, Bonomi AG. Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. IEEE Trans Biomed Eng. 2011; 58(9): 2656–2663.
16. van Hees VT, Golubic R, Ekelund U, Brage S. Impact of study design on development and evaluation of an activity-type classifier. J Appl Physiol. 2013; 114(8): 1042–1051.
17. Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Estimation of free-living energy expenditure using a novel activity monitor designed to minimize obtrusiveness. Obes (Silver Spring). 2010; 18(9): 1845–1851.
18. Schulze M, Song B, Gietzelt M, Wolf K-H, Kayser R, Tegtbur U, et al. Supporting rehabilitation training of COPD patients through multivariate sensor-based monitoring and autonomous control using a Bayesian network: prototype and results of a feasibility study. Inform Health Soc Care. 2010; 35(3–4): 144–156.
19. Faurholt-Jepsen D, Hansen KB, van Hees VT, Christensen LB, Girma T, Friis H, et al. Children treated for severe acute malnutrition experience a rapid increase in physical activity a few days after admission. J Pediatr. 2014; 164(6): 1421–1424.
20. Ortlieb S, Dias A, Gorzelniak L, Nowak D, Karrasch S, Peters A, et al. Exploring patterns of accelerometry-assessed physical activity in elderly people. Int J Behav Nutr Phys Act. 2014; 11(1): 28.
21. Morseth B, Ahmed LA, Bjørnerem Å, Emaus N, Jacobsen BK, Joakimsen R, et al. Leisure time physical activity and risk of non-vertebral fracture in men and women aged 55 years and older: the Tromsø Study. Eur J Epidemiol. 2012; 27(6): 463–471.
22. da Silva IC, van Hees VT, Ramires V V, Knuth AG, Bielemann RM, Ekelund U, et al. Physical activity levels in three Brazilian birth cohorts as assessed with raw triaxial wrist accelerometry. Int J Epidemiol. 2014; 43(6): 1959–1968.
23. Sabia S, van Hees VT, Shipley MJ, Trenell MI, Hagger-Johnson G, Elbaz A, et al. Association Between Questionnaire- and Accelerometer-Assessed Physical Activity: The Role of Sociodemographic Factors. Am J Epidemiol. 2014; 179(6): 781–790.
24. Ortlieb S, Gorzelniak L, Nowak D, Strobl R, Grill E, Thorand B, et al. Associations between multiple accelerometry-assessed physical activity parameters and selected health outcomes in elderly people – results from the KORA-age study. PLoS One. 2014; 9(11): e111206.
25. Emaus A, Degerstrøm J, Wilsgaard T, Hansen BH, Dieli-Conwright CM, Furberg A-S, et al. Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromso study. Scand J Public Health. 2010; 38(5 Suppl): 105–118.

26. van der Berg JD, Stehouwer CDA, Bosma H, van der Velde JHPM, Willems PJB, Savelberg HHCM, et al. Associations of total amount and patterns of sedentary behaviour with type 2 diabetes and the metabolic syndrome: The Maastricht Study. Diabetologia. 2016; 59(4): 709–718.
27. Troiano R, Mc Clain J. Objective measures of physical activity, sleep, and strength in U.S. National Health and Nutrition Examination Survey (NHANES) 2011–2014. In: Proceedings of the 8th International Conference on Diet and Activity Methods. Rome, Italy; 2012.
28. Knuth AG, Assunção MCF, Gonçalves H, Menezes AMB, Santos IS, Barros AJD, et al. [Methodological description of accelerometry for measuring physical activity in the 1993 and 2004 Pelotas (Brazil) birth cohorts]. Cad Saude Publica. 2013; 29: 557–565. Portuguese.
29. Ried-Larsen M, Brønd JC, Brage S, Hansen BH, Grydeland M, Andersen LB, et al. Mechanical and free living comparisons of four generations of the Actigraph activity monitor. Int J Behav Nutr Phys Act. 2012; 9: 113.
30. Rowlands A V, Stone MR, Eston RG. Influence of speed and step frequency during walking and running on motion sensor output. Med Sci Sport Exerc. 2007; 39(4): 716–727.
31. Brage S, Wedderkopp N, Franks PW, Andersen LB, Froberg K. Reexamination of validity and reliability of the CSA monitor in walking and running. Med Sci Sports Exerc. 2003; 35(8): 1447–1454.
32. Brage S, Brage N, Wedderkopp N, Froberg K. Reliability and Validity of the Computer Science and Applications Accelerometer in a Mechanical Setting. Meas Phys Educ Exerc Sci. 2003; 7(2): 101–119.
33. Godfrey A, Barry G, Mathers JC, Rochester L. A comparison of methods to detect postural transitions using a single tri-axial accelerometer. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2014. p. 6234–6237.
34. Lukowicz P, Junker H, Troster G. Automatic Calibration of Body Worn Acceleration Sensors. Lect Notes Comput Sci. 2004; (3001): 176–181.
35. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva IC, et al. Auto-calibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. J Appl Physiol (1985). 2014; 117(7): 738–744.
36. Bai J, He B, Shou H, Zipunnikov V, Glass TA, Crainiceanu CM. Normalization and extraction of interpretable metrics from raw accelerometry data. Biostatistics. 2014; 15(1): 102–116.
37. Brandes M, VAN Hees VT, Hannöver V, Brage S. Estimating energy expenditure from raw accelerometry in three types of locomotion. Med Sci Sports Exerc. 2012; 44(11): 2235–2242.
38. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sport Exerc. 2008; 40(1): 181–188.
39. Tudor-Locke C, Brashear MM, Johnson WD, Katzmarzyk PT. Accelerometer profiles of physical activity and inactivity in normal weight, overweight, and obese U.S. men and women. Int J Behav Nutr Phys Act. 2010; 7: 60.
40. German National Cohort (GNC) Consortium. The German National Cohort: aims, study design and organization. Eur J Epidemiol. 2014; 29(5): 371–382.
41. Sabia S, Cogranne P, van Hees VT, Bell J a., Elbaz A, Kivimaki M, et al. Physical Activity and Adiposity Markers at Older Ages: Accelerometer Vs Questionnaire Data. J Am Med Dir Assoc. 2015; 16(5): 438.e7–13.
42. Shannon CE. Communication in the Presence of Noise (Classic Paper). Proceedings-IEEE. 1998; 86(2): 447–457.
43. Welch WA, Bassett DR, Thompson DL, Freedson PS, Staudenmayer JW, John D, et al. Classification accuracy of the wrist-worn gravity estimator of normal everyday activity accelerometer. Med Sci Sports Exerc. 2013; 45(10): 2012–2019.
44. Rowlands A V, Fraysse F, Catt M, Stiles VH, Stanley RM, Eston RG, et al. Comparability of Measured Acceleration from Accelerometry-Based Activity Monitors. Med Sci Sports Exerc. 2015; 47(1): 201–210.
45. Dannecker KL, Sazonova NA, Melanson EL, Sazonov ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. Med Sci Sports Exerc. 2013; 45(11): 2105–2112.
46. Vroege DP, Wijsman CA, Broekhuizen K, de Craen AJM, van Heemst D, van der Ouderaa FJG, et al. Dose-response effects of a Web-based physical activity program on body composition and metabolic health in inactive older adults: additional analyses of a randomized controlled trial. J Med Internet Res. 2014; 16(12): e265.
47. Zhang S, Rowlands A V, Murray P, Hurst TL. Physical activity classification using the GENEA wrist-worn accelerometer. Med Sci Sports Exerc. 2012; 44(4): 742–748.
48. Stiles VH, Griew PJ, Rowlands A V. Use of accelerometry to classify activity beneficial to bone in premenopausal women. Med Sci Sports Exerc. 2013; 45(12): 2353–2361.
49. Bonomi AG, Plasqui G. "Divide and conquer": assessing energy expenditure following physical activity type classification. J Appl Physiol. 2012; 112(5): 932; author reply 933.
50. Brazendale K, Beets MW, Bornstein DB, Moore JB, Pate RR, Weaver RG, et al. Equating accelerometer estimates among youth: The Rosetta Stone 2. J Sci Med Sport. 2016; 19(3): 242–249.
51. Bornstein DB, Beets MW, Byun W, Welk G, Bottai M, Dowda M, et al. Equating accelerometer estimates of moderate-to-vigorous physical activity: in search of the Rosetta Stone. J Sci Med Sport. 2011; 14(5): 404–410.
52. Kelly P, Fitzsimons C, Baker G. Should we reframe how we think about physical activity and sedentary behaviour measurement? Validity and reliability reconsidered. Int J Behav Nutr Phys Act. 2016; 13(1): 32.
53. Hickey AM, Freedson PS. Utility of Consumer Physical Activity Trackers as an Intervention Tool in Cardiovascular Disease Prevention and Treatment. Prog Cardiovasc Dis. 2016; 58(6): 613–619.
54. Liu S, Gao RX, Freedson PS. Computational methods for estimating energy expenditure in human physical activities. Med Sci Sports Exerc. 2012; 44(11): 2138–2146.