Establishing a Global Standard for Wearable Devices in Sport and Fitness: Perspectives from the New England Chapter of the American College of Sports Medicine Members

Garrett I. Ash, PhD[1,2], Matthew Stults-Kolehmainen, PhD[3,4], Michael A. Busa, PhD[5], Robert Gregory, PhD[6], Carol Ewing Garber, PhD, FACSM[4], Jason Liu, BS[7], Mark Gerstein, PhD[7,8], José Antonio Casajus, MD[9,10], Alex Gonzalez-Aguero, PhD[9,10], Demitri Constantinou, MD[10,11], Michael Geistlinger, PhD, Jur[10,12], Fergus M. Guppy, PhD[13], Fabio Pigozzi, MD[10,14], Yannis P. Pitsiladis, PhD, FACSM[10,14-15]

Garrett.ash@yale.edu
Matthew.stults-kolehmainen@ynhh.org
mbusa@umass.edu
gregoryr3@southernct.edu
garber@tc.columbia.edu
jason.j.liu@bulldogs.yale.edu
mark@gersteinlab.org
joseant@unizar.es
alexgonz@unizar.es
Demitri.Constantinou@wits.ac.za
Michael.Geistlinger@sbg.ac.at
F.Guppy@brighton.ac.uk
fabio.pigozzi@uniroma4.it
Y.Pitsiladis@brighton.ac.uk

[1]*Veterans Affairs Connecticut Healthcare System, West Haven, Connecticut, USA*
[2]*Center for Medical Informatics, Yale University, New Haven, Connecticut, USA*
[3]*Bariatric Surgery Program, Yale-New Haven Hospital, New Haven, Connecticut, USA*
[4]*Department of Biobehavioral Sciences, Teachers College, Columbia University, New York, New York, USA*
[5]*Institute for Applied Life Sciences, University of Massachusetts Amherst, Amherst, Massachusetts, USA*
[6]*Department of Health and Movement Sciences, Southern Connecticut State University, New Haven, Connecticut, USA*
[7]*Department of Molecular Biophysics and Biochemistry, Yale University*
[8]*Department of Computer Science, Yale University*
[9]*Growth, Exercise, Nutrition and Development (GENUD) Research Group, University of Zaragoza, Zaragoza, Spain*
[10]*International Federation of Sports Medicine (FIMS), Lausanne, Switzerland*
[11]*Centre for Exercise Science and Sports Medicine, FIMS Collaborating Centre of Sports Medicine, School of Therapeutic Sciences, Faculty of Health Sciences, University of the Witwatersrand, South Africa*
[12]*University of Salzburg, Salzburg, Austria*
[13]*Centre for Stress and Age-related Disease, School of Pharmacy and Biomolecular Sciences (PaBS), University of Brighton, Brighton, United Kingdom*

[14]*Department of Movement, Human and Health Sciences, University of Rome "Foro Italico", Rome, Italy*
[15]*Collaborating Centre of Sports Medicine, University of Brighton, Eastbourne, United Kingdom*

Corresponding Author:
Yannis P. Pitsiladis
Collaborating Centre of Sports Medicine
University of Brighton
Eastbourne
United Kingdom
Email: y.pitsiladis@brighton.ac.uk

Funding:
GA is supported by a fellowship grant from the Office of Academic Affiliations at the United States Department of Veterans Affairs.

Conflict of Interest: The authors have no conflict of interest to report.

ABSTRACT

The recent explosion of wearable technology and the associated concerns prompted the International Federation of Sports Medicine (FIMS) to create a quality assurance standard for Wearable Devices, which provides commissioned testing of marketing claims and endorsement of commercial wearables that test favorably. An Open Forum as announced in the conference advertising was held at the annual meeting of the New England Regional Chapter of the American College of Sports Medicine (NEACSM) on November 7-8, 2019 in Providence, Rhode Island, USA for attending NEACSM members to voice their input on the process. Herein we report the proceedings. The round table participants perceived the quality assurance standard to be important, but identified some practical process challenges that included the broad scope and complexity of the device universe, the need for a multiphase testing pathway, and the associated fees for product evaluation. The participants also supported the evaluation of device data analysis, behavioral influences, user experience in the overall evaluation. Looking forward, the

FIMS quality assurance standard faces the challenge of balancing these broader perspectives with practical constraints of budget, facilities, time and human resource.

KEYWORDS: connected devices; information technology; sports; physical fitness; global wearable standard.

INTRODUCTION

Wearable technology, the top worldwide fitness trend since 2016, is advancing rapidly and unpredictably (1-3). Wearables projected to be a $27 billion industry by 2022 (1-3). Defined as technology worn on or close to the body these devices can assess position, motion, impact, biomechanical forces, heart rate, muscle oxygen saturation, and/or sleep patterns (4). Examples include the Fitbit® watch, the Mamori® mouthguard sensing head trauma, the CricFlex® elbow sleeve sensing arm forces, the Polar® chest strap measuring heart rate, the Humon Hex® thigh strap optically sensing oxygenated hemoglobin, and the Huawei® watch analyzing wrist movement and heart rate to produce sleep statistics. A more comprehensive overview can be seen in Figure 1. Briefly, the left contains the medically certified devices, the right the sports and fitness trackers, and the middle the "secondary healthcare market" (products outside the scope of medical certification yet typically having high accuracy, reliability, and accessibility).

Figure 1. Overview on the wearable market as defined by Wearable Technologies (http://www.wearable-technologies.com).

Each device possesses unique advantages and challenges. The United States Food and Drug Administration (FDA) regulates wearable devices used to diagnose or treat health conditions, however, no equivalent regulation exists for consumer sport and fitness trackers.

The International Federation of Sports Medicine (FIMS) aims to promote the well-being of all who are engaged in sports and exercise and assist athletes in achieving optimal performance. In this context, the Scientific Commission of FIMS advocated for a global standard for sport and fitness wearables amidst the rising concerns for quality assurance related to the products (3). Such concerns have been rising for several years. After reviewing 61 consumer wearables in 2018, Peake et al. reported that only 5% matched the marketing claims based on accepted reference standards (5). One 2016 study found inaccurate heart rates and consumers sued Fitbit® (6). Wearable devices threaten individual privacy and lack the security afforded

most personal data. For example, data obtained from these devices do not fall into purview of laws governing health data privacy, consequently workplace wellness programs can provide fitness tracker data to insurers, who may, in turn, raise premiums of individuals with high-risk (*e.g.*, sedentary) behavior patterns (7). These threats to individual data exist even when companies claim the data is de-identified, because 24hr biodata has a unique signature, similar to DNA (8). Finally, the best practice for data interpretation and presentation to consumers remains controversial. For instance, there have been reports of wearable sleep watches causing a *"preoccupation or concern with improving or perfecting wearable sleep data"* to the point of ignoring medical advice, not following standard sleep hygiene education, and ignoring laboratory sleep device data with greater validity (9).

In January 2019, FIMS established a task force to address the need for a quality assurance standard for wearable devices. The task force recommended that wearables manufacturers pay a testing fee to enable testing of the company's marketing claims and ensure the device has the advertised capabilities relative to the current gold standard research measurement tool or an appropriate proxy. For instance, the global positioning system (GPS) precision for devices tracking football players would be higher than those tracking marathon runners. While no current regulations exist requiring companies to complete this testing, validating claims may provide a competitive advantage to companies completing testing and provide a product quality endorsement.

Stakeholder meetings began in January 2019, and the first FIMS Collaborating Centre of Sports Medicine (FIMS CCSM) (GENUD Research Group, University of Zaragoza, Zaragoza, Spain) hosted the initiative in September 2019. The multidisciplinary GENUD Lab has a record of accomplishment for designing and implementing interventions that combine a nutritional-

physical activity-psychological approach. GENUD has experts in in assessing body composition and indirect calorimetry and has a long-standing record of performing clinical and public health investigations together with nurses, dietitians, and physical activity trainers. GENUD also has extensive experience with method validation focusing on body composition and physical activity assessment in children, adolescents, and adults.

Wearable Technologies (http://www.wearable-technologies.com), an innovation and market development platform supporting wearables manufacturers is a key partner in the FIMS Global Wearable Standard. This partnership will appoint future testing centers and establish inter-center reliability to prevent bias from single centers. The testing results will be submitted to peer-reviewed journals for publication. FIMS will organize global meetings to disseminate the wearable standard findings and endorsements. GENUD is establishing protocols and standard operating procedures for testing and certification of wearable devices, with the next steps to include internal validations (January 2020 – February 2020), testing (March 2020 – June 2020) then officially certifying the first devices (July 2020) before full implementation (August 2020).

The feedback from New England Chapter of the American College of Sports Medicine (NEACSM) members on the FIMS wearables quality assurance standard is featured in this report. ACSM was a strategic partner in the initial phases of developing the FIMS quality assurance standard. To obtain timely ACSM input to the process, the NEACSM November 2019 (Providence, RI) meeting was utilized to sample the ACSM membership perspectives.

"The New Guiding Reference Standard for Wearable Devices by the International Federation of Sports Medicine: Open Forum for ACSM Membership Feedback" with authors GA and MS as moderators was open to all interested members. Personal invitations were sent electronically to wearables experts from the regional chapter membership and these experts were

requested to invite their colleagues and students to attract a broad and diverse a range of opinions. Thirty-one chapter members attended the Forum including 9 higher education faculty, 2 physical therapists, 1 registered dietitian, 3 military researchers, and 16 students.

GA presented the background information for the FIMS quality assurance standard during the first 20 min, and then led a 30 min discussion based on questions determined prior to the session. Fourteen attendees contributed to the discussion with both verbal and written comments. An audio recording of the session was transcribed, combined with the written comments, and summarized by author GA. The three most active participants (MB, RG, CG) were invited to edit the summary and clarify their responses. The discussion included data science (*e.g.*, wearables using Artificial Intelligence) and two scientists from the Yale University Department of Molecular Biophysics and Biochemistry with specific experience analyzing wearable sensor data (JL and MG) were recruited to develop the relevant sections.

NEACSM ROUNDTABLE THEMES

The scope of the problem is broad

There was general agreement among NEACSM participants that scope of wearables present a broad problem including: 1) the number and types of devices, 2) the reliability, validity, and sensitivity evaluation metrics that need to be tested for each device (10), 3) the context in which the same type of device needs to be evaluated (*e.g.*, measuring impact of a golf club is different than measuring impact of a tennis racket), and 4) the populations in which each device will need to be evaluated (*e.g.*, accuracy of photoplethysmography heart rate measurements taken by Fitbit® may vary according to skin color) (11). The device software is frequently updated, and the Food and Drug Administration has established a regulatory framework for

modifications to software that fall under its jurisdiction (12). Considering all of these variables, undertaking all of this at once may be an "insurmountable" task.

There are several potential strategies for managing the roll out of the FIMS quality assurance standard. First, establish a best practice guideline for each type of device before moving forward with testing. The best existing example is for accelerometers (discussed below), although a more granular approach might be needed to account for the different contexts (*e.g.*, acceleration while running is different than walking or jumping). Other societies may already have established best practice guidelines (*e.g.*, the American Society of Biomechanics) that could be applied to the FIMS quality assurance standard. Second, consider restricting the focus of the FIMS quality assurance standard to a narrow area (*e.g.*, accelerometers tracking locomotion) to establish general best practices before expanding to other areas. Third, establish a multi-year phase in plan rather than arbitrarily imposing a completion date of 2020. Fourth, consider a *"tiered"* testing approach that breaks up the specific components (*e.g.*, validity versus reliability, laboratory versus free-living) rather than attempting to test all variables at once. There is merit to an organization such as FIMS overseeing the global standard for wearable devices in sport and fitness, namely the non-profit nature of the organization that is bound to impartiality and the highest ethical standards. FIMS should also explore the opportunity to place the quality assurance standard in the FDA pathway.

<u>A multiphase pathway is needed</u>

Keadle et al. developed a "Framework to Evaluate Devices that Assess Physical Behavior" (13), which includes five phases: mechanical signal testing, laboratory development, semi-structured evaluation, naturalistic validation, and adoption (Figure). Although this model

was developed for testing accelerometers, it should be applicable to testing other types of devices including ergonomics, human factors, and metabolic evaluation (13). For example, while Figure 2 describes the case of accelerometers, it could be alternatively tailored to the case of energy expenditure sensors.

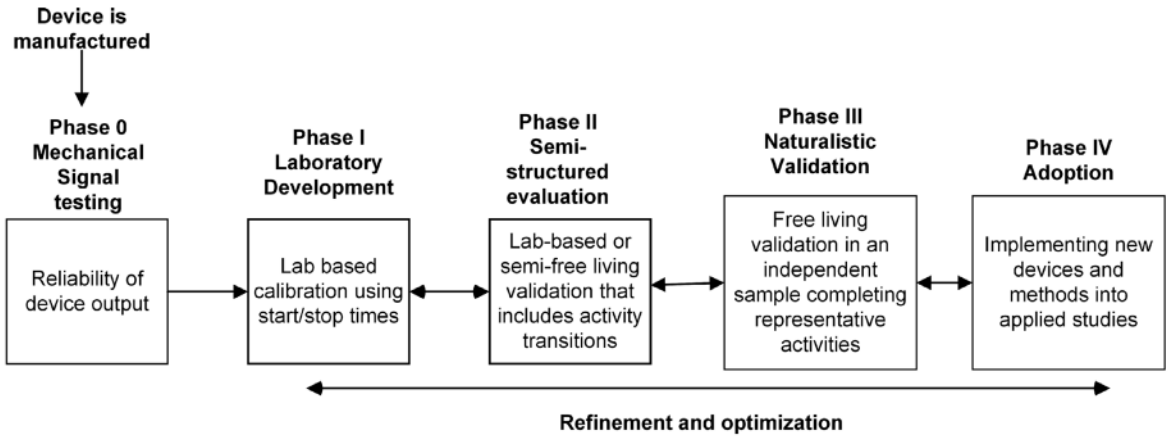

Figure 2. Overview of a phase-based framework to evaluate body-worn devices to assess physical behavior. Reused with permission from (13).

The budget requires reevaluation

The budget for evaluation of 10,000€ per device proposed by FIMS is likely inadequate. A commercial device tested for three different metrics (movement, behavior, and sleep) utilizing methods outlined by Keadle et al. that involved 20 healthy young individuals required more than 300,000USD total. A portion of this study was conducted in a simulated domicile with a kitchen, bathroom, bedroom, and living room. To validate behavioral classifiers, participants stayed in the domicile for 48 hours while simultaneous second-by-second video and sensor data were collected; yielding 960 hours of video that required approximately 4000 person-hours to

annotate. Clearly the proposed 10,000€ budget would be insufficient to cover the costs of this evaluation.

Some potential cost-savings mechanisms were suggested. First, devices that share core pieces of hardware (*e.g.*, different models of the Fitbit may share the same inertial measurement unit) would only need phase 0 testing to be completed once, although different form factors of devices should be evaluated separately. Second, using common core facilities could save overhead costs. For instance, the University of Massachusetts Institute for Applied Life Sciences has *"core facilities for projects ranging from device prototyping, precision manufacturing and roll-to-roll fabrication, to human motion and gait studies, calorimetry, magnetic resonance imaging and spectroscopy, as well as, EEG and sleep studies. These facilities are equally accessible to academic, government, and industry collaborators"* (14) and using these facilities would reduce the cost of duplicating the facilities. Third, collaborating on refining testing pathways can reduce redundancy and avoid repeating inefficiencies. For instance, in the above example, a lower number of observations may have been adequate and extra steps can be eliminated when testing devices in the future. Fourth, manufacturers should be made aware of the lawsuit penalties (*e.g.*, Fitbit class action lawsuit) that the FIMS quality assurance testing process could avert as incentive to pay the true cost of testing fees. Fifth, habitual physical activity is now recognized as an outcome variable in some industry-sponsored drug trials so sport and fitness devices should be explored for application to medical contexts that may attract medical companies as third-party stakeholder contributors (although this would raise the stakes for FDA approval). There are currently 167 active trials on clinicaltrials.gov using a Fitbit® (15). Furthermore, the United States FDA has certified mobile health applications (Pear Therapeutics,

Boston, USA) through the Investigational New Drug pathway opening the door for other digital interventions to be classified this way.

Go beyond the data: data analysis, behavioral influence, and user experience

Wearables collect and internally process data, in some cases utilizing artificial intelligence. For example, mobile applications can provide insulin adjustment advice to people with diabetes based on their continuous glucose monitor readings (16). The data must be presented in a way that allows for useful inference for patients and prevents information overload that may confound good intentions. Since real-world data may not follow fixed theoretical distributions, data interpretation may require statistical models or numerical methods that rely on estimation, which may introduce errors into the output. To ensure the model fits the data and is easy to adapt and interpret, the FIMS quality assurance standard should include data scientists in the evaluation process.

The Forum also raised concerns that clients might over-interpret the data in a way that may hinder attention or motivation. For instance, similar to the preoccupation with wearable sleep device data mentioned in the introduction, athletes could become preoccupied with heart rate variability data from the Whoop® device leading them to ignore coaching advice about training load.

User experience is an important consideration in addition to the analytic data and user errors could impact data quality. These errors could be nuanced such as an improperly positioned watch, or errors of omission such as forgetting to press the start button; YP noted this happens ~50% of the time among elite marathoners (Personal communication YP). The privacy and sharing issues raised are concerning in an era of computer hacking and potential adverse use by

companies. Lastly, there are safety concerns that involve chemical batteries, such as a smartwatch that was recalled due to overheating and, in some cases, burning users' wrists (17). Wearable Technologies (together with TÜV SÜD, Germany) has developed consumer safety certification of wearable devices, so that area is of concern is viewed as outside the scope of the FIMS quality assurance standard.

Poll of priorities

At the end of the session, we asked attendees to complete a poll assigning a 1 to 4 priority score to each category of sports wearables (Table), however the data scientists involved in this forum would reverse this order, based on the extent to which an error could compromise health and safety. The data scientists also would be likely to prioritize products that pose either high medical liability due to being surgically implanted or high security liability due to using cloud storage that may be accessible by others.

Table. Poll results. 20 of the 31 attendees participated.

Median Priority Score	Category of Device
#1	Position and motion
#2	Heart rate and muscle oxygen saturation
#3	Biomechanics
#4	Impact

CONCLUSIONS

The FIMS quality assurance standard is perceived to be very important, but the forum identified practical challenges including the broad scope and complexity of the problem, the need for a multiphase pathway, and unrealistic cost estimates. In addition, the forum supported data analysis, behavioral influence, and user experience as priorities for the quality assurance standard. The forum participants from sports and exercise medicine placed greater priority on devices that are most commonly used (*i.e.*, position and heart rate sensors), whereas those from data science placed greater priority on devices where an error could compromise health and safety (*i.e.*, impact sensors). Looking toward the future, the FIMS quality assurance standard will face the challenge of balancing these broader perspectives with practical constraints.

ACKNOWLEGMENTS

The authors thank the organizers of the conference, especially Dr. Chee-hoi Leong for facilitating logistics of the roundtable and Dr. David Korfhagen for transcribing the session recording.

REFERENCES

(1) Success of apple watch means more growth in sales of wearable technology [Internet]. CCS Insight; [cited 2019 November 20]. Available from: https://www.ccsinsight.com/press/company-news/3695-success-of-apple-watch-means-more-growth-in-sales-of-wearable-technology/.

(2) Thompson WR. Worldwide survey of fitness trends for 2019. *ACSMs Health Fit J.* 2019;23 (6):10-8.

(3) Duking P, Stammel C, Sperlich B, et al. Necessary steps to accelerate the integration of wearable sensors into recreation and competitive sports. *Curr Sports Med Rep* 2018; Jun;17(6):178-82.

(4) Seshadri DR, Li RT, Voos JE, et al. Wearable sensors for monitoring the internal and external workload of the athlete. *NPJ Digit Med* 2019:1-18.

(5) Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. *Front Physiol* 2018; Jun 28;9:743.

(6) Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. *J Sports Sci Med* 2016; Aug 5;15(3):540-7.

(7) Wellness programs raise privacy concerns over health data [Internet]. Alexandria (VA): The Society for Human Resource Management; [cited 2019 November 20]. Available from: https://www.shrm.org/resourcesandtools/hr-topics/technology/pages/wellness-programs-raise-privacy-concerns-over-health-data.aspx.

(8) Peraklsis E, Coravos A. Is health-care data the new blood? *Lancet Dig Health* 2019; May;1:e8-e9.

(9) Baron KG, Abbott S, Jao N, Manalo N, Mullen R. Orthosomnia: Are some patients taking the quantified self too far? *J Clin Sleep Med.* 2017; Feb 15;13(2):351-4.

(10) Duking P, Fuss FK, Holmberg HC, Sperlich B. Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. *JMIR Mhealth Uhealth* 2018; Apr 30;6(4):e102.

(11) Shcherbina A, Mattsson CM, Waggott D, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *J Pers Med* 2017; May 24;7(2):10.3390/jpm7020003.

(12) U.S. Food and Drug Administration. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning-Based Software as a Medical Device [Internet]. White

Oak (MD): U.S. Food and Drug Administration; [cited 2019 November 21]. Available from: https://www.fda.gov/media/122535/download.

(13) Keadle SK, Lyden KA, Strath SJ, Staudenmayer JW, Freedson PS. A framework to evaluate devices that assess physical behavior. *Exerc Sport Sci Rev* 2019; Oct;47(4):206-14.

(14) University of Massachusetts Amherst. Institute for Applied Life Sciences [Internet]. Amherst (MA): University of Massachusetts Amherst; [cited 2019 November 21]. Available from: https://www.umass.edu/ials/about.

(15) Clinicaltrials.gov Search Results [Internet]. Bethesda (MD): U.S. National Library of Medicine; [cited 2019 November 29]. Available from: https://clinicaltrials.gov/ct2/results?term=fitbit&Search=Apply&recrs=a&age_v=&gndr=&type=&rslt=.

(16) Breton MD, Patek SD, Lv D, et al. Continuous glucose monitoring and insulin informed advisory system with automated titration and dosing of insulin reduces glucose variability in type 1 diabetes mellitus. *Diabetes Technol Ther* 2018; Aug;20(8):531-40.

(17) CNN. Burns from fitness trackers prompt recall of about 30,000 gadgets [Internet]. Atlanta (GA): CNN; [cited 2019 November 21]. Available from: https://www.cnn.com/2018/09/26/health/provata-fitness-tracker-recall/index.html.