www.ispor.org

ISPOR

Improving healthcare decisions

Workshop 6: Using Performance Outcome Assessments in the Evaluation of Clinical Benefit in Multinational Treatment Trials: Unique Challenges and Considerations Along the Path from Selection to Implementation of a Fit-for-Purpose Measure

Presented by the ISPOR Performance Outcome Assessment Emerging Good Practices Task Force

ISPOR Europe

November 13, 2023

Antitrust Compliance Statement

- ISPOR has a policy of strict compliance with both United States, and other applicable international antitrust laws and regulations.
- Antitrust laws prohibit competitors from engaging in actions that could result in an unreasonable restraint of trade.
- ISPOR members (and others attending ISPOR meetings and/or events) must avoid discussing certain topics when they are together including, prices, fees, rates, profit margins, or other terms or conditions of sale.
- Members (and others attending ISPOR meetings and/or events) have an obligation to terminate any discussion, seek legal counsel's advice, or, if necessary, terminate any meeting if the discussion might be construed to raise antitrust risks.
- The Antitrust policy is available on the ISPOR website at ispor.org/antitrust.

Disclaimer

- The views and opinions expressed in the following slides are those of the individual presenters and should not be attributed to their respective organizations/companies or the U.S. Food and Drug Administration.
-

Funding Acknowledgment

- Critical Path Institute (C-Path) is supported by the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) and is 55% funded by FDA/HHS, totaling $17,612,250, and 45% funded by non-government source(s), totaling $14,203,111. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS or the U.S. Government. For more information, please visit FDA.gov

Moderator:

- **Sonya Eremenco, MA,** Executive Director, Patient-Reported Outcome (PRO) Consortium, Critical Path Institute, Tucson, AZ, USA

Speakers:

- **Rachel Ballinger, PhD,** Principal, ICON, Reading, England, UK
- **Bill Byrom, PhD,** Principal eCOA Science, Signant Health, Nottingham, England, UK
- **Michelle Campbell, PhD,** Associate Director, Stakeholder Engagement and Clinical Outcomes, Office of Neuroscience, FDA, Silver Spring, MD, USA

PerfO Assessment Task Force Leadership Group Report 1

- **Heather R. Adams, PhD**, Pediatric Neuropsychologist and Associate Professor, University of Rochester, USA
- **Rachel Ballinger, PhD, BSc,** Principal, Patient Centred Outcomes, ICON, UK
- **Elizabeth (Nicki) Bush, MHS (Task Force Co-Chair),** Senior Director, Endpoints and Measurement Strategy, Johnson and Johnson, USA
- **Bill Byrom, PhD**, Principal eCOA Science, Signant Health, UK
- **Michelle Campbell, PhD,** Associate Director, Stakeholder Engagement and Clinical Outcomes, Office of Neuroscience, FDA, Silver Spring, MD, USA
- **Wen-Hung Chen, PhD,** Director, Patient-Centered Outcomes, GSK, USA
- **Stephen Joel Coons, PhD, (Task Force Co-Chair),** Senior Vice President, Critical Path Institute, USA

PerfO Assessment Task Force Leadership Group Report 1 (cont.)

- **Helen Doll, Dphil,** Senior Principal, Clinical Outcome Solutions, UK
- **Chris Edgar, PhD**, **(Task Force Co-Chair),** Chief Science Officer, Cogstate, London, England, UK
- **Sonya Eremenco, MA**, Executive Director, PRO Consortium, Critical Path Institute, USA
- **Fiona McDougall, PhD**, Senior Director, PRO Research, Genentech, USA
- **Elektra Papadopoulos, MD, MPH,** Director, Patient Experience Data & Strategy in Immunology and Oncology, AbbVie, USA
- **Bray Patrick-Lake, MFS,** Senior Digital Health Specialist, Digital Health CoE, CDRH, FDA, USA
- **Ashley F. Slagle, PhD,** Principal, Scientific and Regulatory Consulting, Aspen Consulting, LLC, USA

PerfO Assessment Task Force Leadership Group Report 2

- **Heather R. Adams, PhD**, Pediatric Neuropsychologist and Associate Professor, University of Rochester, USA
- **Alexandra Atkins, PhD,** Sr. Director, Value, Evidence and Outcomes Research, Neuroscience, Eli Lilly & Company, USA
- **Rachel Ballinger, PhD, BSc,** Principal, Patient Centred Outcomes, ICON, UK
- **Elizabeth (Nicki) Bush, MHS (Task Force Co-Chair),** Senior Director, Endpoints and Measurement Strategy, Johnson and Johnson, USA
- **Bill Byrom, PhD**, Principal eCOA Science, Signant Health, UK
- **Michelle Campbell, PhD,** Associate Director, Stakeholder Engagement and Clinical Outcomes, Office of Neuroscience, FDA, Silver Spring, MD, USA
- **Heather Dickerson, BA,** Senior Outcomes Researcher, Patient Centred Outcomes, ICON, USA

PerfO Assessment Task Force Leadership Group Report 2 (cont.)

- **Chris Edgar, PhD**, **(Task Force Co-Chair),** Chief Science Officer, Cogstate, London, England, UK
- **Sonya Eremenco, MA**, Executive Director, PRO Consortium, Critical Path Institute, USA
- **Naomi Knoble, PhD**, Associate Director, Rare Disease Measurement Science, DCOA, FDA, USA
- **Fiona McDougall, PhD**, Senior Director, PRO Research, Genentech, USA
- **Ashley F. Slagle, PhD,** Principal, Scientific and Regulatory Consulting, Aspen Consulting, LLC, USA
- **Kevin Weinfurt, PhD**, Vice Chair for Research, Department of Population Health Sciences, Duke University, USA

Outline

1. Introduction to Performance Outcome (PerfO) Assessment Task Force and Overview of PerfO Assessments| Sonya Eremenco

2. Implementation of PerfO assessments in multinational clinical trials | Rachel Ballinger

3. Use of digital health technology (DHT) to administer PerfO assessments and/or collect PerfO data | Bill Byrom

4. Regulatory expectations regarding fit-for-purpose PerfO assessments in the evaluation of clinical benefit | Michelle Campbell

5. Questions, Answers, Discussion| Moderated by Sonya Eremenco

Definition: Performance Outcome (PerfO) Assessment

A type of clinical outcome assessment (COA). A measurement based on standardized task(s) actively undertaken by a patient according to a set of instructions. A PerfO assessment may be administered by an appropriately trained individual or completed by the patient independently. Examples of PerfO assessments include:

- Measures of gait speed (e.g., timed 25-foot walk test using a stopwatch or using sensors on ankles)
- Measures of memory (e.g., word recall test)

Source: FDA-NIH Biomarkers, EndpointS, and other Tools (BEST) Resource Glossary

PerfO Assessments in the Context of Other COAs

Four COA types:

- Patient-reported outcome (PRO) measures
- Clinician-reported outcome (ClinRO) measures
- Observer-reported outcome (ObsRO) measures
- PerfO assessments

COAs assess clinical benefit: how a patient feels, functions, or survives.

Regulatory evidentiary expectations are similar across COAs supporting endpoints in clinical trials

Source: FDA-NIH Biomarkers, EndpointS, and other Tools (BEST) Resource Glossary

How PerfO Assessments Differ from Other COA Types

- A PerfO assessment is used when the optimal means of capturing the clinical benefit of therapeutic interventions is through the physical completion of defined/standardized tasks that reflect or are the foundational building blocks for day-to-day activities that are important and meaningful in patients' lives.
- Primary relevant aspects of function: cognitive, mobility, and sensory
- Often used in pediatric or cognitively impaired populations where assessment via the other COAs is not possible
- Relationship between the concept of interest (COI) and the meaningful aspect of health may be indirect rather than direct as with other COAs
- Requires standardization of administration
- Able to address heterogeneity in target population better than other COAs

Potential Limitations with Other COA Types

1	Concept(s) of interest not best reported by the patient (e.g., memory impairment)
2	Difficulty observing concept(s) of interest without prompting task performance (e.g., tremor in movement disorders not evident at rest, but posture, movement, or task dependent)
3	Issues with recall periods (e.g., COI(s) infrequently, rarely performed and/or assessment of current state needed)
4	Patient biased in reporting/rating leading to under or over-estimation (e.g., bias due to negative affect or loss of insight)
5	Observer biased in reporting/rating leading to under or over-estimation (e.g., bias due to psychosocial factors/relationship with patient)

Task Force Scope

The task force reports will address PerfO assessments of physical function (e.g., mobility), cognitive function (e.g., working memory) or cognition-dependent function (e.g., instrumental activities of daily living), and sensory function (e.g., low contrast visual acuity).

Task Force Goal

To enhance the appropriate use and usefulness of PerfO assessments in the evaluation of clinical benefit in medical product development by providing consensus-driven good practice recommendations regarding:

- The development, selection, and modification of PerfO assessments, including the evaluation and documentation of validity, reliability, usability, and interpretability (Report 1); and

- The scientific and operational issues associated with appropriate and effective PerfO assessment implementation in clinical trials (Report 2 to be discussed today).

ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

ISPOR Report

Recommendations on the Selection, Development, and Modification of Performance Outcome Assessments: A Good Practices Report of an ISPOR Task Force

Chris J. Edgar, PhD, Elizabeth (Nicki) Bush, MHS, Heather R. Adams, PhD, Rachel Ballinger, PhD, BSc, Bill Byrom, PhD, Michelle Campbell, PhD, Sonya Eremenco, MA, Fiona McDougall, PhD, Elektra Papadopoulos, MD, MPH, Ashley F. Slagle, PhD, Stephen Joel Coons, PhD

ABSTRACT

In evaluating the clinical benefit of new therapeutic interventions, it is critical that the treatment outcomes assessed reflect aspects of health that are clinically important and meaningful to patients. Performance outcome (PerfO) assessments are measurements based on standardized tasks actively undertaken by a patient that reflect physical, cognitive, sensory, and other functional skills that bring meaning to people's lives. PerfO assessments can have substantial value as drug development tools when the concepts of interest being measured best suit task performance and in cases where patients may be limited in their capacity for self-report. In their development, selection, and modification, including the evaluation and documentation of validity, reliability, usability, and interpretability, the good practice recommendations established for other clinical outcome assessment types should continue to be followed, with concept elicitation as a critical foundation. In

Edgar CJ, Bush EN, Adams HR, et al. Recommendations on the selection, development, and modification of performance outcome assessments: a good practices report of an ISPOR Task Force. *Value Health*. 2023;26(7):959–967.

Task Force Report 1 Summary

- PerfO assessment development and validation follows the same foundational good practice as for other COA types
  - Concept elicitation is central to understanding what to measure
  - Involves validation of assessment and its score and interpretation of the score
- To identify or develop tasks to support PerfO assessments, the use of both qualitative and quantitative (task performance) data may be of value
  - Motor function assessments
  - Neuropsychological test batteries
  - Sensory function tests
- A relationship between a task or group of tasks and the patient's usual life must still be established in order that the PerfO assessment can evaluate treatment benefit
  - In a unitary validity model, this may be supported using ecological validity evidence

Task Force Report 2: The Work Continues…

- While Report 1 focused on generating evidence that PerfO assessment is fit-for-purpose, Report 2 will focus on implementation in clinical trials to generate valid and reliable data to support endpoints evaluating treatment benefit
- Aspects to be addressed today
  - Challenges in multinational clinical trials
  - Standardizing administration of PerfO assessments across sites in a trial
  - Use of DHT to administer PerfO assessments and/or collect PerfO data
- The report will address these and other topics in our regulatory context keeping recent draft guidance documents in mind

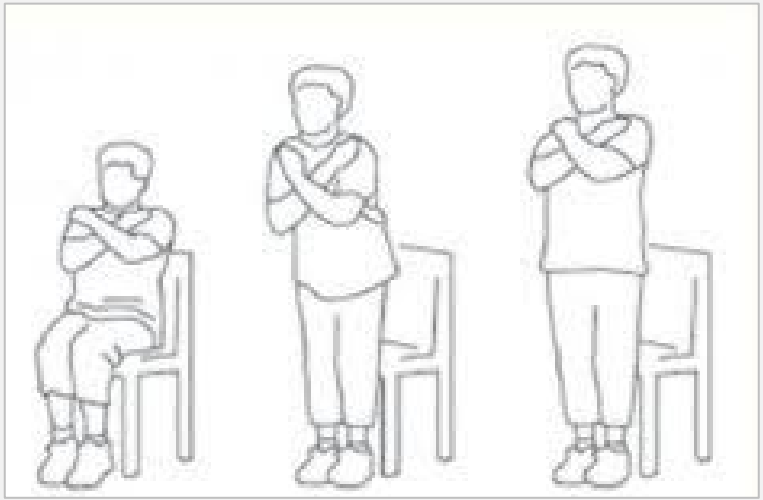
SECTION

1

Implementation of Performance Outcome Assessments in Multinational Clinical trials

Rachel Ballinger, PhD
Principal, Patient Centred Outcomes
ICON plc

Exercise: Repeated Chair Stand



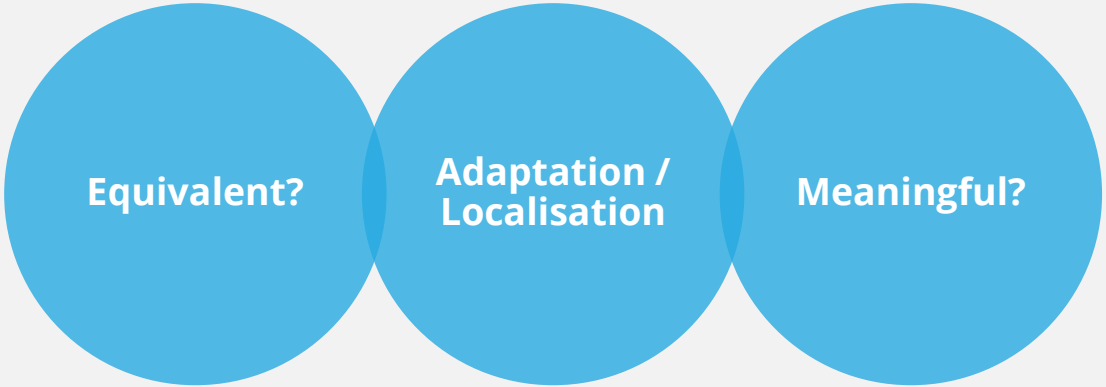
Repeat 5x, as quickly as you can while remaining safe

Please only attempt if you feel safe to do so!

Images: Sites.duke.edu; Nutritionletter.tufts.edu

Cultural Validation and Comparability

"A measurement based on standardized task(s) actively undertaken by a patient according to a set of instructions"



In International Contexts, Day-To-Day Sitting Positions Vary

Does this matter?

Should we:

1. Standardise to have one sitting position?
2. Use different approaches in each location?
3. Use different approaches and take this into account in scoring algorithms?

Raichlen et al. Proc Natl Ada Sci. (2020)

Cross-Cultural Adaptation: Verbal Memory Test

Word Lists Used in the Four Experiments in the Present Study

English List/ USA-English Speakers	English List/ Australian-English Speakers	French List/ French Speakers	French List Translated From Australian/French Speakers	Mandarin List/Brunei Mandarin Chinese Speakers	Malay List/ Brunei Malay Speakers
Fudge	Pavlova	Café	Pavlova	咖啡	Kopi
Brownies	Flake	Pommes	Flocon	苹果	Epal
Candy	Lollies	Steak	Sucettes	牛肉	Daging
Bagel	Pastie	Saumon	Pastie	鲔鱼	Ikan
Pretzel	Lasagne	Oignons	Lasagne	蘑菇	Jagung
Ketchup	Vegemite	Chocolat	Vegemite	薯片	Keropok
Pepperoni	Dim Sim	Champignons	Faible Sim	面粉	Nasi
Pickle	Flathead	Vinaigre	À tête plate	胡椒	Cili
Oatmeal	Weetabix	Crêpes	Weetabix	沙爹	Satay
Soda	Mandarin	Fromage	Mandarin	鸡蛋	Telur
Cornbread	Crumpets	Bonbon	Crumpets	人蔘	Ambuyat
Syrup	Potato Cake	Poivre	Gâteau de pomme de terre	豆奶	Madu

Note—Word lists for each country are not direct translations, but have resulted from the protocol detailed in the Appendix.

Lim, Y.Y., et al. Behavior Research Methods (2009)

Localisation: Activities in Daily Living (digital administration)

The individual must complete a series of tasks related to making a meal. They include:

- Read a recipe, search the pantry at home.
- Make a shopping list of the missing items.
- Take the correct bus to the store.
- Pay the fare in exact change.
- Locate the items on the shopping list in the store.
- Pay for the purchases with exact change.
- Take the correct bus home.

United States

United Kingdom

Bus stop : Bus stop moved to other side of street; Localization of bus design, street signs, background cars, street names

https://isctm.org/public_access/Autumn2016/Poster/Pdfs/Atkins-Autumn-2016-Poster.pdf

Levels of Localisation

Task Presentation

- Different presentation, but the concepts measured are the same
- Task comprehension, e.g., digital administration

Implementation

- Awareness of different environments
- Physical environment e.g., floor coverings for walking assessments
- Equipment set up, tech support, connectivity
- Practice rounds
- Data capture, including models and any battery power

Score Interpretation

- Raw scores unchanged
- Develop algorithms or adjusted scores to account for variation e.g., number of steps and landing steps in site's available stair wells used for long stair climb assessments

Doll et al., Value in Health (2018)

Intent of localisation is to standardise the tasks and interpretation

Helping to Ensure High Quality and Reliable Data

"A PerfO assessment may be administered by an appropriately trained individual or completed by the patient independently"

Site and Location Feasibility

Sufficient physical space for equipment and assessment

- E.g., 6-minute walk test to be performed in a "minimally trafficked area along a flat, straight corridor ideally ≥ 30 m in length to be consistent with established reference equations"
  Gibbons et al, J Cardiopulm Rehab (2021)
- Assess at the time the assessments are likely to take place (not when conveniently quiet!)

Patient safety

- Includes any indemnification needs (especially at sites): "a type of agreement wherein one party agrees not to hold another party liable for legal causes of action in the future"[1]
- Independent settings: safety assessment
  - In general, does the study require patient to do more or anything differently (e.g., faster) than they would otherwise do in daily life? Might any self-administration inadvertently encourage this?
  - Individual patient, e.g., assess suitability of their home setting

Patient privacy

Administrator Training

- **Are there established trainings (modules, materials) for the selected PerfO assessment? Check with the instrument developers**

- **Attestation of eligibility/ qualification & certification to administer the selected PerfO assessments in a specific study**

Robust training is a better indicator of standardized performance than experience
Targum, J Clin Psychopharmacol. (2006)

Ongoing training:
Re-certification and replacement administrator trainings

Unique for PerfO assessments:
Ability to train and support patients for independent completions (i.e., self-administration)

Example Training Content

Background
- Trial context
- What the PerfO assessment is measuring
- How the PerfO assessment was developed and has been used

Study specific details
- PerfO assessment time points in the schedule of administration
- How they are being assessed (in-person, remote)
- Setting up and any ongoing calibration of the equipment assessments

Preparation of the assessment
- Materials needed and set up procedures (in-person, remote)
- Preparing the patient: rapport, ensure they are rested and as comfortable as possible
- Considerations of timing, and administration to atypical patients, e.g., behavioural manifestations

Variations of the training
- Differentiate training for novice vs. advanced administrators
- Available in local languages, video instruction

Assessment administration
- Demonstration and explanation of the task
- Use of instructional script
- Encouragement: may/not be scripted
- Spatial considerations (potential support vs. not impacting ability)
- Patient safety as the top priority, discontinue the assessment if necessary
- Do not: rush the patient, provide feedback, assist them with the task
- Remote monitoring of patient's self-administration

Data collection
- Counting, timing. (note: clinical judgment for scoring = ClinRO)
- Documentation

Supporting materials
- Manuals, scripts, printed materials
- Modules with quiz elements
- Skills demonstration, video reviews
- Opportunities for remediation/retaking
- Ongoing access to training modules for administrators
- Ongoing support resources
- Patient materials for independent completions and 24/7 support

Scoring and Surveillance

Minimal or absent judgment means that **there is no true "rater"** for PerfO assessments

Scoring could be a simple count if the PerfO assessment is measuring the number of correct outcomes, or the use of a scoring algorithm for more complex assessments

Scoring surveillance analyses should include:

- Intra-administrator reliability
- Inter-administrator reliability
- Inter-administrator reliability between the two administrators (e.g., site and remote) across the study time points
- Performance of the administrator, including patient's completing the assessments independently
- Detection of outliers
- Quality checks when using any scoring algorithms

If differences are detected by these regular analyses, it is critical to investigate and provide administrator remediation as early as possible

Conclusion

The goal is to **collect valid data, in a standardized manner**, across **every site, study team, and administration**

Application of difference to help reach standardization

Appreciation of difference to help reach standardization

Unique considerations of administrator role, and patients' self-administration

SECTION

2

Use of Digital Health Technology (DHT) to Administer PerfO Assessments and/or Collect PerfO Data

Bill Byrom, PhD
Principal eCOA Science, Signant Health,
Nottingham, England, UK

DHT-Administration Examples

1. Using a tablet to time and enable manual entry of the recorded distance walked during an in-clinic six-minute walk test (6MWT)
2. Using a video consultation to assess a simple range of motion PerfO remotely
3. Using a sensor/app combination to fully time and record an in-clinic timed-up-and-go test
4. Using a sensor/app combination to enable a patient to complete an at-home postural stability test as part of their weekly COA assessments, including PROMs

The Promise of DHTs

A PerfO is:

- A type of clinical outcome assessment.
- A measurement based on standardized task(s) actively undertaken by a patient according to a set of instructions.
- A PerfO assessment may be administered by an appropriately trained individual **or completed by the patient independently**.
- PerfO assessments include:
  - Measures of gait speed (e.g., timed 25-foot walk test using a stopwatch or using sensors on ankles);
  - Measures of memory (e.g., word recall test)

The Promise of DHTs

DHTs have potential to:

1. Improve measurement
   - Measure something more accurately
   - Measure something with greater detail (e.g., include partial sit/stand transitions)
   - Measure something more conveniently

2. Measure a concept that couldn't be measured practically before

3. Measure a concept more frequently (inc. in other settings, e.g., at home)

Administering a PerfO Assessment Using a DHT

1. Migration (and/or modification) of an existing PerfO assessment to measure using a DHT

2. Development of a new DHT-derived PerfO assessment

Migration of a PerfO Assessment to DHT Measurement

Considerations

1. What evidence is needed to support the suitability of the sensor solution used to instrument the test?

2. What evidence is needed to demonstrate measurement comparability?

3. Is the PerfO-based endpoint sensor-specific?

4. Using sensor data for richer insights: development of additional complementary endpoint measures

Measurement Comparability

VALUE IN HEALTH 21 (2018) 581–589

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

PATIENT-REPORTED OUTCOMES

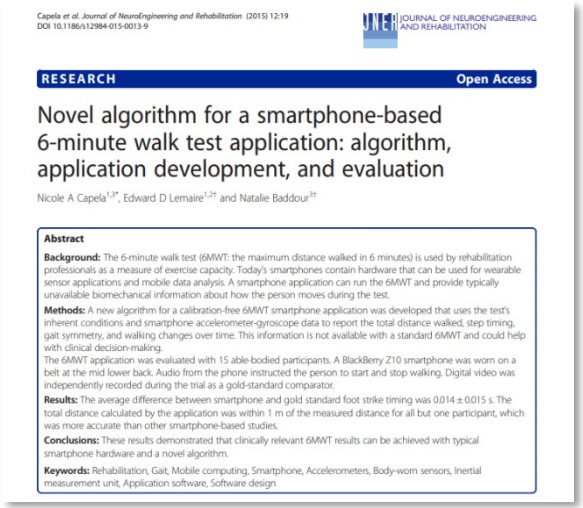
Measurement Equivalence of Patient-Reported Outcome Measure Response Scale Types Collected Using Bring Your Own Device Compared to Paper and a Provisioned Device: Results of a Randomized Equivalence Trial

Bill Byrom, PhD[1,*], Helen Doll, DPhil[1], Willie Muehlhausen, DVM[1], Emuella Flood, BA[1], Cater Cassedy, MA[1], Bryan McDowell, MBA[2], Jeremy Sohn, BA[2], Kyle Hogan[3], Ryan Belmont, MBA[3], Barbara Skerritt[1], Marie McCarthy, MBA[1]

[1]ICON Clinical Research, USA, UK and Ireland; [2]Novartis Pharmaceuticals, Basel, Switzerland; [3]Clinical Ink, Winston-Salem, NC, USA

Measurement Comparability

Example 1. In-clinic 6-minute walking test



Capela et al. Journal of NeuroEngineering and Rehabilitation (2015) 12:19
DOI 10.1186/s12984-015-0013-9

JOURNAL OF NEUROENGINEERING AND REHABILITATION

RESEARCH Open Access

Novel algorithm for a smartphone-based 6-minute walk test application: algorithm, application development, and evaluation

Nicole A Capela[1,3*], Edward D Lemaire[1,2†] and Natalie Baddour[3†]

Abstract

Background: The 6-minute walk test (6MWT: the maximum distance walked in 6 minutes) is used by rehabilitation professionals as a measure of exercise capacity. Today's smartphones contain hardware that can be used for wearable sensor applications and mobile data analysis. A smartphone application can run the 6MWT and provide typically unavailable biomechanical information about how the person moves during the test.

Methods: A new algorithm for a calibration-free 6MWT smartphone application was developed that uses the test's inherent conditions and smartphone accelerometer-gyroscope data to report the total distance walked, step timing, gait symmetry, and walking changes over time. This information is not available with a standard 6MWT and could help with clinical decision-making.
The 6MWT application was evaluated with 15 able-bodied participants. A BlackBerry Z10 smartphone was worn on a belt at the mid lower back. Audio from the phone instructed the person to start and stop walking. Digital video was independently recorded during the trial as a gold-standard comparator.

Results: The average difference between smartphone and gold standard foot strike timing was 0.014 ± 0.015 s. The total distance calculated by the application was within 1 m of the measured distance for all but one participant, which was more accurate than other smartphone-based studies.

Conclusions: These results demonstrated that clinically relevant 6MWT results can be achieved with typical smartphone hardware and a novel algorithm.

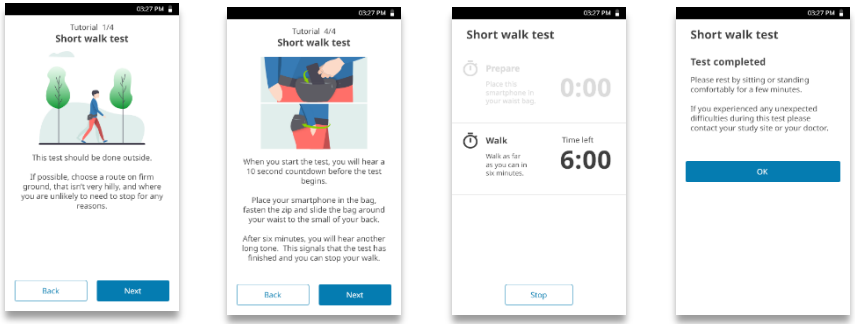
Keywords: Rehabilitation, Gait, Mobile computing, Smartphone, Accelerometers, Body-worn sensors, Inertial measurement unit, Application software, Software design

[1] Capela, N.A. et al. Novel algorithm for a smartphone-based 6-minute walk test application: algorithm, application development, and evaluation. *J NeuroEngineering Rehabil* **12**, 19 (2015).

- DHT – smartphone app
- Algorithm:
  1. Detect number of 180º turns taken
     - Gyroscope, magnetometer
  2. Calculate stride length
     - Number of steps per course length completed
  3. Calculate distance walked
     - Number of course lengths x 25 m
     - + number of steps on final (incomplete) length x stride length
- Mean error: 0.12%
  - True: 542.89 m, DHT: 542.42 m

Measurement Comparability

Example 2. At-home 6-minute walk



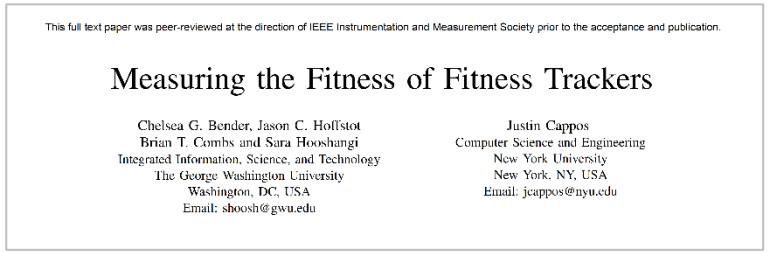
[2] Byrom B. (unpublished).
https://www.biorxiv.org/content/10.1101/2021.10.21.465337v1.full

- DHT – smartphone app
- Algorithm:
  1. GPS
     - Distance travelled
  2. Accelerometer
     - Number of steps
     - Rest intervals
- Mean absolute percentage error:
  - Distance: 1.2 to 1.3%
  - Number of steps: 1.7 to 1.8%

Measurement Comparability

	Traditional approach	In-clinic instrumented approach [1]	At-home instrumented approach [2]
6MWT: 6-minute walking test	Patient asked to walk as far as they can along an in-clinic corridor route of defined length. Accompanied by a trained administrator who provides encouragement every 60s. Distance travelled calculated by the healthcare professional based on the number of lengths completed and use of a measuring wheel for partial length.	Patient asked to walk as far as they can along an in-clinic corridor route of defined length wearing a mobile sensor containing an accelerometer and gyroscope. Accompanied by a trained administrator who provides encouragement every 60s. Distance travelled calculated by the number of lengths completed and the estimated number of steps for partial length.	Patient asked to walk for six minutes outdoors unsupervised using a smartphone app accessing the smartphone in-built accelerometer and a GPS. Distance travelled calculated based on the GPS signal trace.
Concept of interest	Functional capacity: distance walked at maximal pace	Functional capacity: distance walked at maximal pace	Walking performance: distance walked at a comfortable pace

Are Endpoints Sensor-Specific?



This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication.

Measuring the Fitness of Fitness Trackers

Chelsea G. Bender, Jason C. Hoffstot
Brian T. Combs and Sara Hooshangi
Integrated Information, Science, and Technology
The George Washington University
Washington, DC, USA
Email: shoosh@gwu.edu

Justin Cappos
Computer Science and Engineering
New York University
New York, NY, USA
Email: jcappos@nyu.edu

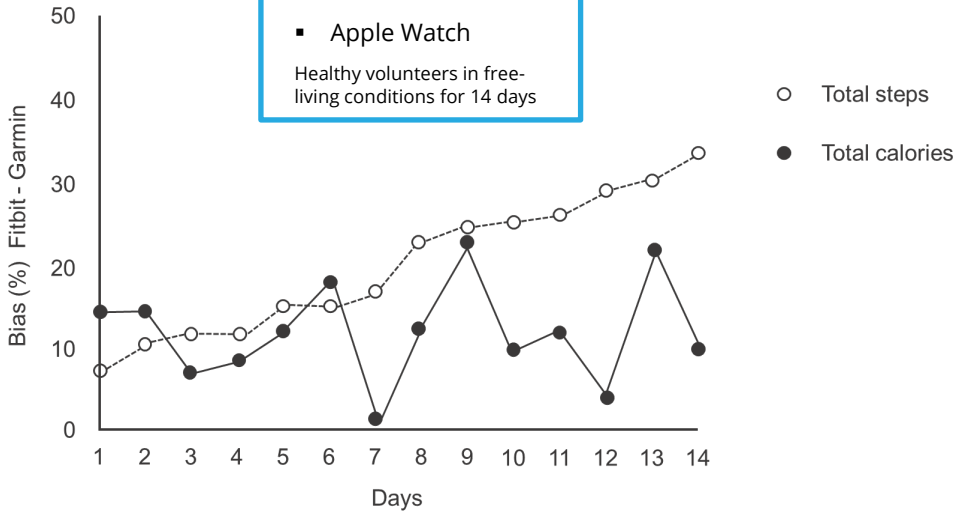
Bender CG, Hoffstot JC, Combs BT et al. Measuring the Fitness of Fitness Trackers. Sensors Applications Symposium 2017 IEEE, pp. 1-6, 2017, March.

> Step count, distance travelled, and calories burned could vary significantly between devices used concurrently.

Devices studied

- Fitbit Flex
- Fitbit Charge HR
- Garmin vivoactive
- Apple Watch

Healthy volunteers in free-living conditions for 14 days



Considerations for Interchanging Devices

- Importance of device-agnostic measures
  - Ability to switch devices mid-study / mid-program
  - Ability to compare / pool the results from multiple studies
  - Ability to compare the results across multiple programs
- Do the devices generate the same measure?
- Do they use the same algorithm?
  - If NOT
    - Can we show evidence that the measures are comparable?
    - Can we access raw sensor data and apply a common algorithm?
- What does comparable mean?
  - What level of tolerance in estimates can we accept as not adversely impacting the overall endpoint measure?

Richer Insights Using Sensors

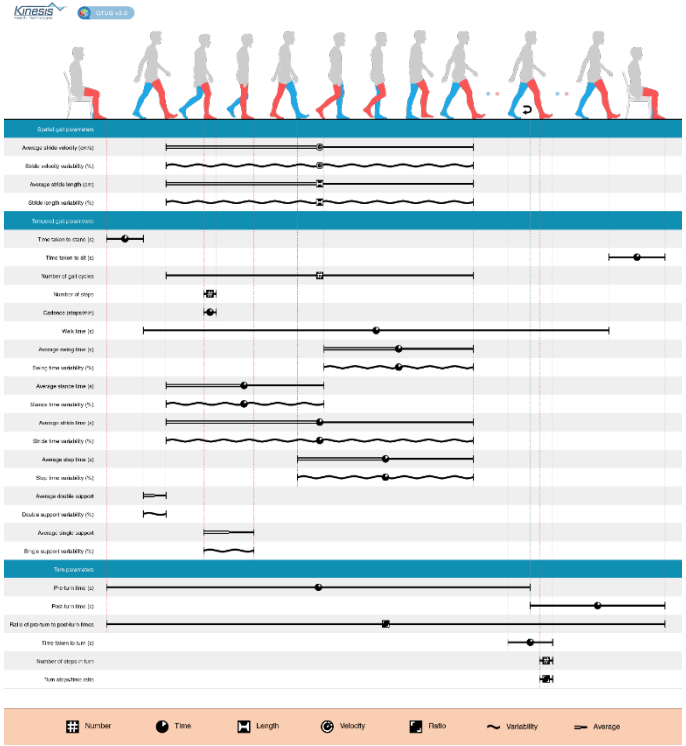


Image © Kinesis Health Technologies Ltd, Dublin, Ireland. Reproduced from: [3] Byrom B. The use of new digital endpoints? Digital endpoints. In (Schueler P, Ed.) *Clinical Trial Methodologies: Lessons Learned during the Corona Pandemic.* Elsevier, 2021.

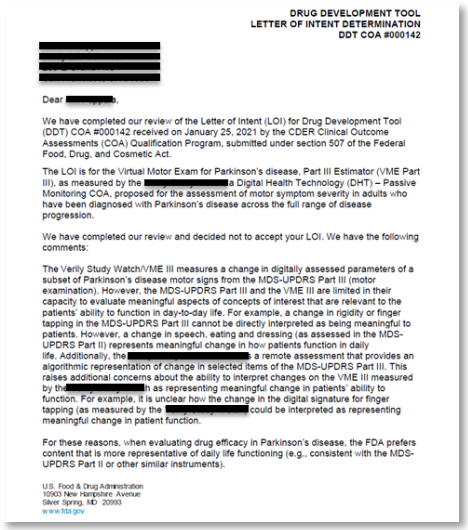
Timed Up and Go (TUG) test

- Inertial sensors on each ankle
- Gait and balance data provide powerful predictors of falls risk [4]
- New PerfO measures require full COA development process

[4] Greene BR et al. Longitudinal assessment of falls in patients with Parkinson's disease using inertial sensors and the Timed Up and Go test. J Rehabil Assist Technol Eng. 2018 Jan 12;5:2055668317750811.

DHT-Administered PerfO Measure Development

DHT endpoint developers do not always follow the COA development framework



DRUG DEVELOPMENT TOOL
LETTER OF INTENT DETERMINATION
DDT COA #000142

[redacted]

Dear [redacted],

We have completed our review of the Letter of Intent (LOI) for Drug Development Tool (DDT) COA #000142 received on January 25, 2021 by the CDER Clinical Outcome Assessments (COA) Qualification Program, submitted under section 507 of the Federal Food, Drug, and Cosmetic Act.

The LOI is for the Virtual Motor Exam for Parkinson's disease, Part III Estimator (VME Part III), as measured by the [redacted] a Digital Health Technology (DHT) – Passive Monitoring COA, proposed for the assessment of motor symptom severity in adults who have been diagnosed with Parkinson's disease across the full range of disease progression.

We have completed our review and decided not to accept your LOI. We have the following comments:

The Verily Study Watch/VME III measures a change in digitally assessed parameters of a subset of Parkinson's disease motor signs from the MDS-UPDRS Part III (motor examination). However, the MDS-UPDRS Part III and the VME III are limited in their capacity to evaluate meaningful aspects of concepts of interest that are relevant to the patients' ability to function in day-to-day life. For example, a change in rigidity or finger tapping in the MDS-UPDRS Part III cannot be directly interpreted as being meaningful to patients. However, a change in speech, eating and dressing (as assessed in the MDS-UPDRS Part II) represents meaningful change in how patients function in daily life. Additionally, the [redacted], a remote assessment that provides an algorithmic representation of change in selected items of the MDS-UPDRS Part III. This raises additional concerns about the ability to interpret changes on the VME III measured by the [redacted] as representing meaningful change in patients' ability to function. For example, it is unclear how the change in the digital signature for finger tapping (as measured by the [redacted] could be interpreted as representing meaningful change in patient function.

For these reasons, when evaluating drug efficacy in Parkinson's disease, the FDA prefers content that is more representative of daily life functioning (e.g., consistent with the MDS-UPDRS Part II or other similar instruments).

U.S. Food & Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993
www.fda.gov

Finger tapping test – Parkinson's Disease

Finger tapping

> "A change in rigidity or finger tapping in the MDS-UPDRS Part III cannot be directly interpreted as being meaningful to patients."

https://www.fda.gov/drugs/clinical-outcome-assessment-coa-qualification-program/clinical-outcome-assessments-coa-qualification-program-submissions

Backfill Example: WATCH-PD

Journal of Parkinson's Disease 13 (2023) 589–607
DOI 10.3233/JPD-225122
IOS Press

Research Report

Mapping Relevance of Digital Measures to Meaningful Symptoms and Impacts in Early Parkinson's Disease

Jennifer R. Mammen[a,*], Rebecca Yang[d], Michelle Campbell[e], Josh Jensen-Roberts[d], Melissa Kostrz Cedarbaum[k,l], E. Ray Dorsey[d,i],
[a]University of Rhode Island, Colleg
[b]Critical Path Institute, Tucson, AZ,
[c]Department of Neurological Scienc
[d]Center for Health + Technology, U
[e]Center for Drug Evaluation and Re
MD, USA
[f]Abbvie Inc., North Chicago, IL, US
[g]Biogen, Cambridge, MA, USA
[h]H. Lundbeck A/S, Valby, Denmark
[i]Department of Neurology, Universi
[j]Northwestern University Feinberg
[k]Coeruleus Clinical Sciences LLC,
[l]Yale Medical School, New Haven, C

Shape rotation task

May map to ability to button/unbutton shirt / other fine motor skill aspects of daily living

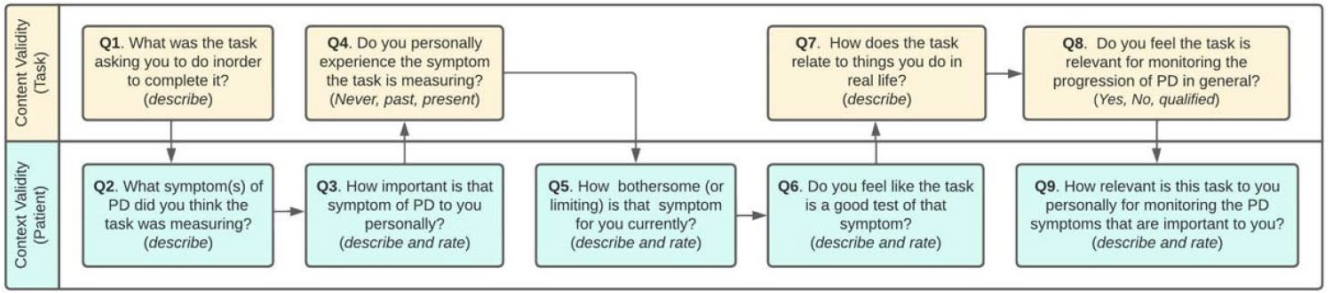


Fig. 5. Recommended approach to assessing the relevance of digital measures for monitoring meaningful symptoms of disease. Use of a consistent 0–10 rating scale for each rated item (i.e., 0 = not important at all; 10 = most important, etc.) could improve comparison across technologies and trials.

Conclusion

- Evidencing measurement comparability is an important aspect of PerfO assessment migration to DHT administration
- Careful consideration of DHT interchangeability is needed when measuring using new DHTs
- DHT-administered PerfO measures need to follow the same COA development process as all COAs
  - Importance of mapping measure to a meaningful aspect of health
  - Concept elicitation in patients the starting point
  - Bring together digital health and COA development experts

SECTION

3

Regulatory expectations regarding fit-for-purpose PerfO assessments in the evaluation of clinical benefit

Michelle Campbell, PhD
Associate Director, Stakeholder Engagement and Clinical Outcomes, Office of Neuroscience, FDA

Reminder: Clinical Benefit for FDA

- A positive clinically meaningful effect of an intervention, i.e., a positive effect on how an individual feels, functions, or survives.
  - How long a patient lives
  - How a patient feels or functions in daily life

- Can be demonstrated as either:
  - A comparative advantage in treatment of the disease or condition; OR
  - A comparative reduction in treatment-related toxicity

Clinical Meaningfulness

- When selecting your performance measure, consideration should be taken that the measure represents clinically meaningful concepts.
- Utilize both qualitative and quantitative evidence to inform decision making.

Examples

- May not provide clinically meaningful information
  - Clinician reporting exam changes of decreased vibratory sense, decreased movement against resistance, or decreased reflexes in arms/hands.
    - Changes may suggest a change in the disease status but do not reflect any impact on patient symptoms or daily functioning.

- Does provide clinically meaningful information
  - Numbness in hands that interferes with the ability to button clothes
  - Weakness in hands that interferes with ability to hold spoon and eat
  - Weakness in arms causing difficulty carrying groceries

Reflections on Report 1

- Aligns with and complements FDA PFDD Guidance Series
- Reflects Roadmap to Patient-Focused Outcome Measurement for Clinical Trials (Figure 2, Guidance 3)
- Aligns with Appendix D (Performance Outcome Measures) in Guidance 3
- Reflects Guidance 4 considerations on measurement properties needed to support clinical trial endpoint

https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical

Threats to Data Quality

Table 2. Factors influencing the need for PerfO assessment.

COA type(s)	Potential challenges
PRO assessment challenges	Substantial cognitive impairment, including loss of insight or language difficulties on the part of the patient
	Pediatric populations eg, availability of age and/or developmentally appropriate measures
	Concept(s) of interest not best known to the patient, including issues with understanding, quantification, and attribution (ie, a good understanding of the cognitive, motor, or sensory issues that may be causing the issue)
	Self-report influenced by other factors (eg, mood influences report in other symptom domains[4])
ClinRO assessment challenges	Substantial cognitive impairment, including loss of insight or language difficulties on the part of the patient
	Difficulty in observing concept(s) of interest without prompting task performance (eg, tremor in movement disorders may not be evident at rest, but rather posture, movement, or task dependent)
ObsRO assessment challenges	Observer biased in reporting leading to under or overestimation[5,6]
	Difficulty in observing concept(s) of interest without prompting task performance (eg, tremor in movement disorders may not be evident at rest, but rather posture, movement, or task dependent)
Challenges with PRO, ClinRO and/or ObsRO assessments	Specific domains may not be measurable for example a change from an asymptomatic to a symptomatic clinical stage of disease with a single assessment tool Change may not be measurable
	Heterogeneity between and within patients in the activities performed in daily life, physical environment, culture, and language

Note: Several ClinRO assessment approaches make use of standardized tasks, the performance of which is then rated using clinical judgment for example, part III of MDS-UPRDS (Motor Examination) in Parkinson disease, or items 8, 9, and 10 of ADAS-Cog (spoken language, language comprehension, and word finding difficulty) in Alzheimer disease. These are not typically described as PerfO assessments as most of the conduct is more consistent with a ClinRO assessment.
ClinRO indicates clinician-reported outcome; MDS-UPRDS, Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale; ObsRO, observer-reported outcome; PerfO, performance outcome; PRO, patient-reported outcome.

- Important to first determine if a PerfO assessment is the most appropriate COA to be used to measure your concept of interest

Edgar CJ, Bush E, Adams HR, et al. Recommendations on the selection, development, and modification of performance outcome assessments: a good practices report of an ISPOR Task Force. *Value Health*. 2023;26(7):959–967.

Threats to Data Quality

Table 3. Considerations in evaluating feasibility of PerfO assessments.

Area of evaluation	Issue	Examples
Appropriateness of task(s)	Tolerability and fatigue	Acceptable time to complete and frequency - can patients tolerate the task duration and effort required and higher frequency assessment?
	Functions other than the concept of interest do not unduly influence scores on the PerfO assessment	eg, motor function issues do not unduly influence cognitive task performance, or motor task performance does not reflect inability to understand requirements because of cognitive impairment
	Appropriateness and range of measurement	eg, walking only suitable in ambulatory patients
		Do tasks allow for the ability to measure improvement and worsening, including an element of challenge, such that a sufficient range of measurement is ensured without important floor or ceiling effects?
	Translatability and cross-cultural adaptation	For a PerfO assessment that includes a meal preparation task of 'making a cooked pudding' the task may be irrelevant for some patients and cultures, hard to translate, or lacking equivalence (eg, issues of usual diet, local foodstuffs, local weights and measures)
	Suitability and ability to standardize setting	eg, location for 6-minute walk test
	Ensuring patient safety	eg, possibility of fall risk during motor assessment
Practical implementation	Availability of standardized manual	
	Availability of standardized training	
	Availability of translations and cultural adaptations	
	Availability and suitability of equipment and stimuli	
	Need for prebaseline training/familiarization	
	Schedule of assessments	Acceptable frequency - is the task suited to proposed frequency of assessment (eg, absence of problematic learning/practice and then ceiling effects)
	Changes to mode of administration	Existing PerfO assessment tasks may need to be revalidated for different modes of administration (eg, paper and pencil task transitioned to computer-administered)
		Existing PerfO assessments may need to be adapted and validated for remote assessment in unsupervised or supervised (eg, telephone, videocall) contexts

PerfO indicates performance outcome.

- If the feasibility of using a PerfO assessment to support a clinical trial endpoint is not adequately evaluated, data quality concerns may result

Edgar CJ, Bush E, Adams HR, et al. Recommendations on the selection, development, and modification of performance outcome assessments: a good practices report of an ISPOR Task Force. *Value Health*. 2023;26(7):959–967.

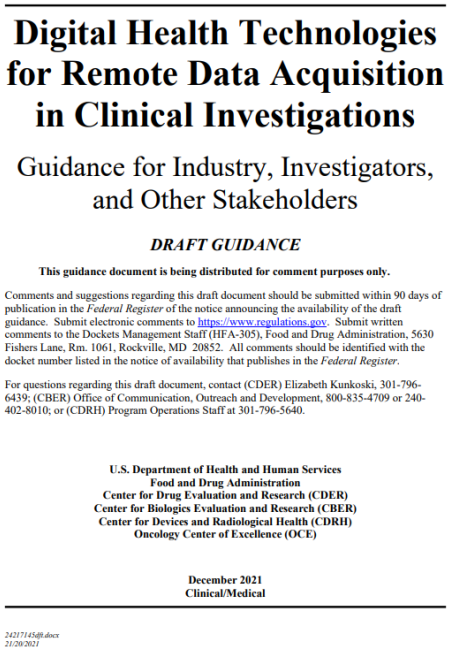
Reflections on Evidence to Support Use of PerfO Assessments

- Is there a good understanding of natural history of the disease and abilities of patient population?
  - Needed to optimize selection of PerfO assessment
- Has the clinical meaningfulness of selected PerfO assessment been established?
  - Example: Does the 6MWT capture a concept that is meaningful to patients and is the change in that distance clinically meaningful to the patient?
- Can we interpret the score?
  - Is the score in raw form? Has the score been transformed? Can I describe the score in labeling clearly?
- Does the PerfO assessment complement other COAs to support overall endpoint hierarchy?
  - What other COAs are included in the trial? Do the other COAs capture concepts the PerfO assessment will not capture?

Global Consideration

- Is selected PerfO assessment applicable to global development programs?
  - Can you implement PerfO assessments in all global trial sites?
- Is additional evidence needed to support use in global trials?
  - Is the concept of interest or meaningful aspect of health the same for all participants in a global development program?
- Considerations should be thought about early in a development program
  - Is translation or cultural adaptation needed?
- FDA does engage with other global health/regulatory agencies

Didn't Forget… DHTs



Digital Health Technologies for Remote Data Acquisition in Clinical Investigations

Guidance for Industry, Investigators, and Other Stakeholders

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Elizabeth Kunkoski, 301-796-6439; (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010; or (CDRH) Program Operations Staff at 301-796-5640.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Oncology Center of Excellence (OCE)

December 2021
Clinical/Medical

2421714Sdft.docx
21/20/2021

- This draft guidance provides recommendations to facilitate the use of DHTs in clinical investigations
- It is designed to help accelerate efficient medical product development to help bring new innovations and advances to patients
- It builds on the launch of the Digital Health Center of Excellence to empower digital health stakeholders and provide regulatory clarity and collaboration across FDA

https://www.fda.gov/media/155022/download

In the End We Want: Fit-for-Purpose Measurement

For medical product development tools, fit-for-purpose is a conclusion that the level of validation associated with a tool is sufficient to support its context of use*

*A statement that fully and clearly describes the way the medical product development tool is to be used and the medical product development-related purpose of the use.

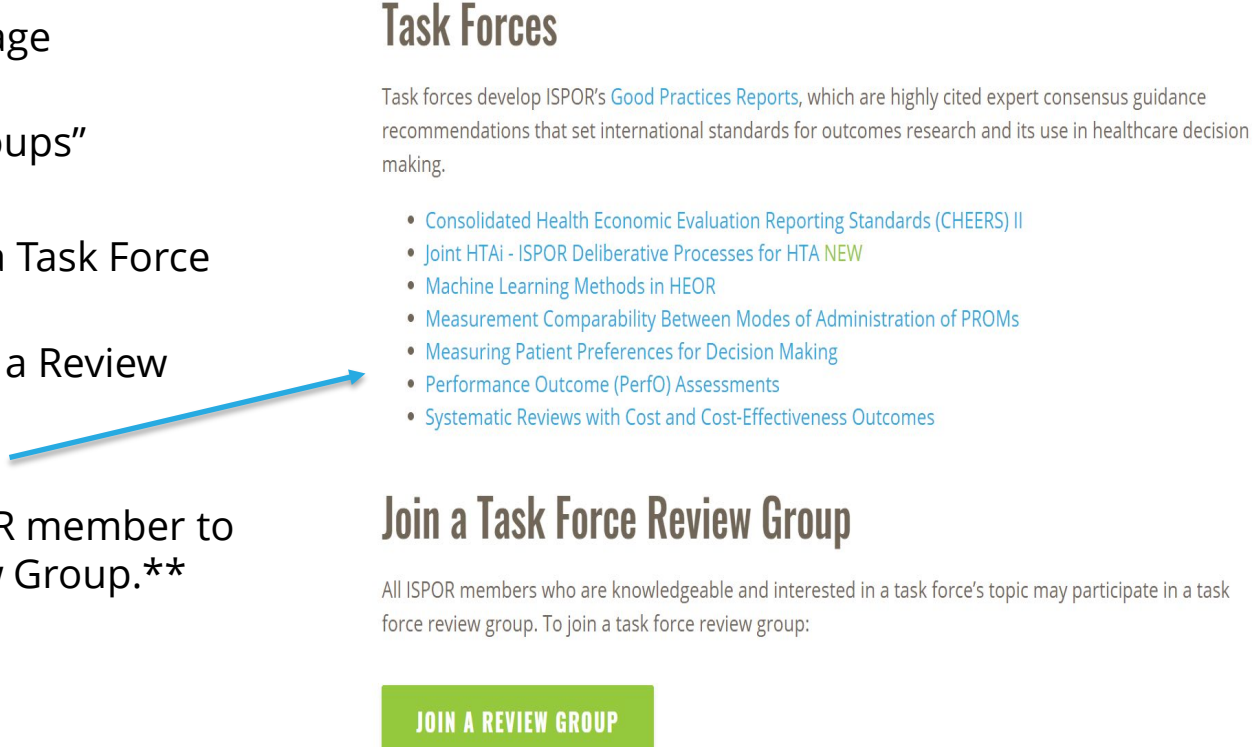
Conclusion

- Know your population
  - Understand if patients can follow instructions to complete assessments
- Standardize Administration of Performance Measures
  - Document (with detail) how measures are to be administered
  - Provide appropriate training for study staff on administration (includes clear and sufficiently detailed user manual)
  - Pilot test to make sure that patients are able to complete the assessment correctly and safely
- Reference available resources, including ISPOR PerfO Task Force Report (note that the report aligns well with the PFDD Guidance Series)
- Come Early for Advice

Join Our Task Force Review Group!

1. Visit ISPOR home page www.ispor.org
2. Select “Member Groups”
3. Select “Task Forces”
4. Scroll down to Join a Task Force Review Group
5. Click button to “Join a Review Group”

You must be an ISPOR member to join a Task Force Review Group.



SECTION

4

Q&A

To contact the presenters:
taskforce@ispor.org

www.ispor.org

ISPOR

Improving healthcare decisions

Related ISPOR Activities

ISPOR Special Interest Groups

- Biosimilars
- Clinical Outcome Assessment (COA)
- Digital Health
- Epidemiology
- Health Preference Research
- Medical Devices & Diagnostics
- Medication Adherence & Persistence
- Nutrition Economics
- Oncology
- Open-Source Models
- Patient-Centered
- Precision Medicine & Advanced Therapies
- Rare Disease
- Real World Evidence (RWE)
- Statistical Methods in HEOR

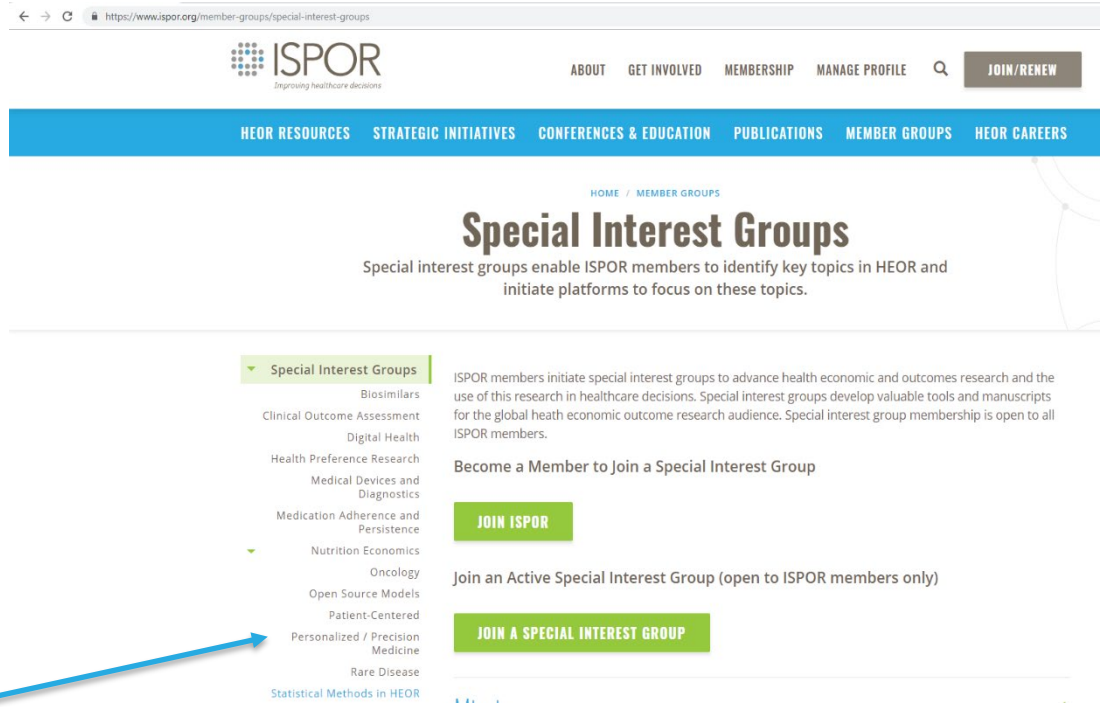
Join an ISPOR Special Interest Group

ISPOR Special Interest Group

1. Visit ISPOR home page www.ispor.org
2. Select “Member Groups”
3. Select “Special Interest Groups”
4. Click button to “Join A Special Interest Group”

For more information, e-mail sigs@ispor.org

You must be an ISPOR member to join a Special Interest Group*




thank you
danke
рахмат
謝謝
ngiyabonga
teşekkür ederim
Баярлалаа
mersi
kia ora
barka
welalin
tack
spas
спасибо
faafetai lava
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
gracias
хвала
asante
manana
obrigada
tenki
kiitos
dankie
nanni
nandri
dhanyavad
maururu
koszonom
dankon
aciu
akun
bedankt
bayarlalaa
gracie
hvala
enkosi
dziękuję
chnorakaloutioun
gratias ago
gracies
sulpay
djiere dieuf
tau
mochchakkeram
murakoze
дякую
mamnun
chokrane
go raibh maith agat
sobodi
dekuji
obrigado
mesi
didi madloba
sagolun
sukriya
kop khun krap
ありがとう
taiku
arigatô
takk
dakujem
trugarez
najis tuke
kam sah hamnida
rahmat
terima kasih
tanemirt
rahmet
grazie
ευχαριστώ
diolch
dhanyavadagalu
shukriya
merce
мерси
감사합니다
xiexie
merci