

Review Article

Validation and equivalency of electronic clinical outcomes assessment systems

Sarah T. Gary*, Antonio V. Otero, Kenneth G. Faulkner, Nadeeka R. Dias

E Research Technology, Inc., Boston, MA, USA

Received: 31 July 2020

Accepted: 11 September 2020

*Correspondence:

Dr. Sarah T. Gary,

E-mail: sarah.gary@ert.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

The US food and drug administration (FDA) has long called for clinical outcomes assessments (COA), such as patient-reported outcomes (PRO), to be ‘fit-for-purpose’ meaning the COA has been validated to support the context of use. The FDA’s recent patient-focused drug development guidance series has renewed the importance of ensuring that COA is ‘fit-for-purpose’ and valid. In addition, the FDA has recommended that COA be collected electronically and that the electronic (eCOA) system and devices also be validated. Advancing technology requires eCOA systems and devices to evolve; eCOA devices may change over time. As bring your own device (BYOD) models gain popularity and acceptance, devices may also be mixed within trials. Changes in eCOA devices or mixing devices may require equivalence testing to prove validity across platforms. The aim of this article is to provide an overview of the different types of validation at both the assessment level and the eCOA system (device) level to help clinical trial sponsors determine the appropriate level of validation or equivalence testing required for COA used in their clinical trials.

Keywords: Validation, Equivalence, BYOD, eCOA, ePRO, Cognitive debriefing/usability testing

INTRODUCTION

COAs, such as PROs are an important part of many clinical trial regulatory submissions, not just to support labelling claims but also for evaluation of the drug’s benefit-risk assessment. Per FDA guidance documents, COAs submitted to the FDA should be ‘fit-for-purpose’ meaning that the COA has been validated sufficiently to support its context of use.^{1,2} In addition, the FDA recommends that COA data should be collected electronically (eCOA), and that the electronic system used to collect the eCOA data undergo a rigorous validation process.^{2,3}

This article describes what it means for a COA to be considered valid, and the difference between COA instrument validation and eCOA system validation. In addition, this article examines how validity is affected when eCOA devices or platforms change or are mixed

within a trial, and what equivalency testing should be performed to maintain validity. The aim of this article is to provide clinical trial sponsors with an overview of what validation or equivalency testing is needed at the assessment (COA) level as well as at the eCOA system level.

OVERVIEW OF TYPES OF VALIDATION

Assessment validation (evaluation of psychometric properties)

The FDA and European medicines agency (EMA) published several guidance documents on the use of COAs in clinical trials, including: FDA ‘Guidance for industry patient reported outcome measures: Use in medical product development to support labelling’, FDA ‘Patient-focused drug development guidance series for enhancing the incorporation of the patient’s voice in

medical product development and regulatory decision making', EMA 'Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products,' and EMA 'Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man-the use of patient-reported outcome (PRO) measures in oncology studies.'^{1,4-6}

The FDA calls for the COA to be 'fit-for-purpose' meaning the COA should be appropriate for the intended use, reliably and validly measure concepts that are clinically relevant and important to the patient population of interest, and understood and interpreted by the user as was originally intended by the COA author. While this document focuses on validity, reliability is also another important measurement property and a requirement for validity.⁷ The following is a summary of evidence recommended to be submitted to the FDA to demonstrate that a COA is 'fit-for-purpose': conceptual framework of the concepts being measured, content validity or evidence that the COA measures the concepts of interest, and evidence of other measurement properties such as reliability, construct validity, and ability to detect change. Content validity is obtained from qualitative studies (e.g. interviews or focus groups), quantitative studies, and/or published literature. Examples of information submitted to establish content validity include: literature review, expert input, concept elicitation interviews, and cognitive interviews to test the COA.

Sponsors may choose to use an existing COA, modify an existing COA, or develop their own COA. In any case, sponsors should show that the COA is 'fit-for-purpose' as outlined above in their patient population. If a sponsor chooses to use an existing COA in a different patient population than the one where it was validated, or modify an existing COA including a change from paper to electronic collection, a small qualitative study in the target patient population may be required, including concept elicitation (CE), and/or cognitive interviews (CI, also called cognitive debriefing or CD) to demonstrate that the COA is still 'fit-for-purpose' in the new patient population or on the new modality.

Mode-to-mode equivalence (e.g. paper-to-electronic equivalence)

As noted above, if an existing COA is administered on a different modality from which it was validated, a sponsor may need to show that the COA is still valid on the new modality. This type of study is often called a mode-to-mode (e.g. Paper-to-Electronic) equivalence study. The purpose of this study is not to show content validity or assess the psychometric properties but rather show that the change in modality does not change the participants' understanding or responses. Typically, equivalence studies are done for COA that will be used to support labelling claims to meet regulatory requirements or to fulfil copyright holder requirements. For example, many

COAs are first developed and validated on paper. To move to a different mode with only minor changes, such as a faithful migration to electronic, the FDA PRO guidance recommends that a small qualitative study may be adequate to show that the assessment still performs as intended on the new modality.¹ Substantial changes may require quantitative equivalence testing or treating the COA as a new instrument with full psychometric testing.⁸

Traditionally, the gold standard for paper-to-electronic equivalence for minor changes was cognitive debriefing and usability testing (CD/UT) as described by Coons et al in the 2009 international society for pharmacoeconomic and outcomes research (ISPOR) task force paper 'Evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures.'⁸ However, recently there has been a shift in the scientific literature showing that an expert screen review may be sufficient in addition to device usability testing in a representative patient population.^{9,10} Recent literature based on CD/UT testing concludes that it is very rare for PROs to be inequivalent when moving from paper to electronic administration, such that an expert screen review and usability testing may be sufficient to demonstrate equivalence.

Linguistic validation

Once a COA is finalized in the source language, it may undergo translations and cultural adaptation. Per the FDA PRO guidance, PRO translations that will support labelling claims require full linguistic validation.¹ The ISPOR task force for translation and cultural adaptation recommends a translation and linguistic validation process that includes dual forward translation and reconciliation, back translation and review, and cognitive debriefing interviews with in-country participants (typically 5 per language) recruited based on the indication, age, and in some cases gender.¹¹ When using eCOA, there are two processes to consider: source document (e.g. paper) translations of the assessment, and then migration of the translated assessment to an electronic system. Linguistic validation can be done at either step.

eCOA system validation

eCOA system validation is a process performed prior to a clinical trial to ensure that the eCOA device and software programming function as intended and demonstrate performance stability for use in a clinical trial.^{2,3} The eCOA system validation process has been outlined in the ISPOR task force paper: 'Validation of electronic systems to collect PRO data-recommendations for clinical trial teams.'³ Generally, the process includes the following eight components:

System requirements definition: System requirements describe all aspects of the system, such as the needs of

the protocol, patients, and clinic staff, regardless of the technology.

Defining system requirements before the eCOA system development begins allows for quick, easy and inexpensive changes to be made; as well as allows the eCOA team to get a clear understanding of the requirements before designing the system.

System design: System design documentation includes detailed and technical description of the eCOA system. System design documentation includes details of the data collection and storage, web portal including display of source data in reports, alerts, and data transfer to the sponsor or CRO.

Coding/software development: Writing code or customizing code that has already been developed in a software programming language to align with the system requirements and system design documentation.

Testing by the eCOA system provider: Testing the eCOA system, as outlined in a test plan and comprehensive set of test cases, to ensure that all aspects of the eCOA system meets the agreed upon system requirements and system design documentation.

Traceability: A traceability matrix that ensures that each system requirement has been accounted for in the system design document as well as in the test cases.

User acceptance testing (UAT): UAT is a process by which the clinical trial team tests the eCOA system and determines whether it meets their expectations and conforms to the system requirements.

Installation and configuration management: Process by which the eCOA provider installs the tested software on the eCOA device, and ensures the correct study version is deployed to the correct location.

Decommissioning plan: Process by which the eCOA provider retires or decommissions the eCOA system when a clinical trial ends. It includes ensuring that all collected patient data are uploaded to the central database and the database is locked, collection of eCOA devices, ensuring that all required validation documentation exists in the vendor archive, and notifying all internal and external support parties.

The eCOA vendor should also provide evidence that the eCOA system is in compliance with FDA regulation 21 CFR Part 11, with regards to electronic records and electronic signatures.¹² It is important to note that eCOA system validation refers specifically to the eCOA device used (computer, smartphone, tablet) and software programming (which includes all aspects of collecting eCOA data, storing eCOA data, displaying eCOA data in reports, and transferring data to sponsors), while assessment validation relates only to the COA itself (the

content of the COA) as previously described. Ideally, assessment validation is performed prior to eCOA system validation to avoid rework of the eCOA system if the COA must be modified as a result of the assessment validation process.

CD/UT VS EXPERT SCREEN REVIEW

Cognitive debriefing/usability testing (CD/UT) is a qualitative research process used to determine whether concepts and items are understood by study participants in the way the instrument developer intended. If electronic data collection is used, usability testing of the device to determine ease of use is performed at the same time as the cognitive interviews. CD/UT typically employs a semi-structured interview with a mix of close-ended questions and a 'think aloud' process.

The typical process for CD/UT with electronic implementation is outlined below: Development of study documents (e.g. Protocol, semi-structured interview guide, informed consent form, case report form etc.) and device programming are performed in parallel, IRB submission of study documents and screenshots, Participant recruitment (usually n=10), Participant interviews (cognitive debriefing and usability testing), Coding and analysis of qualitative data and final report.

CD is typically performed after an instrument has been developed to test the instrument in the patient population of interest and ensure that the participants understand the instrument as intended. In addition, as mentioned previously, CD/UT can also be used to show paper-to-electronic equivalence.⁸

Expert screen review in lieu of CD for paper-to-electronic equivalency was originally proposed by Muehlhausen et al in 2018.⁹ The authors conducted a meta-analysis of 53 CD/UT studies and identified minor findings in only 9% of studies. Taken together with previous quantitative studies that found a high level of agreement between paper and electronic assessments, expert screen review was recommended by the authors instead of CD/UT when migrating an instrument which was originally validated on paper to an electronic form.^{13,14} However, regulatory agencies have yet to formally adopt this recommendation, although guidance updates from the FDA, as well as updated recommendations from ISPOR, are expected in 2020 which may address the topic. Expert screen review can be used to demonstrate that eCOA design best practices have been followed and includes review of the following components by an eCOA expert: review of the electronic assessment instructions, ensuring a faithful representation of the original instrument instructions. In general, there should be minimal changes to instruction text, with only text specific to paper changed (e.g. changing 'circle the response' to 'select the response'), Usability including readability of the font size as represented on the electronic screen and ease of navigation between questions and Item-by-Item

migration review. This includes ensuring no changes in the core wording of the item stem and response options; key emphasis is maintained (e.g. bolding or underlining); for visual and numeric response scales, anchor text should be located so that it is clear which position on the scale is being referenced.

Expert screen review should be coupled with usability testing in the study participant population. Muehlhausen et al have suggested that usability testing in a representative patient population, as opposed to each target patient population, may be sufficient.⁹ The representative patient population should have similar age ranges, education level, and socio-economic status to the target patient population, and may include additional groups such as dexterity-challenged participants, technology-naïve participants, cognitively-challenged participants, or partially-sighted participants.

When to use CD/UT

The COA will be used to support a labelling claim, or be used as part of a submission to regulatory agencies and the COA or parts of the COA have never been tested to ensure understanding in the study participant population. Note that additional validation may be required to show content validity (such as concept elicitation) and other measurement properties as noted previously. Typically, CD/UT should be performed well prior to phase 3 studies (Figure 1).

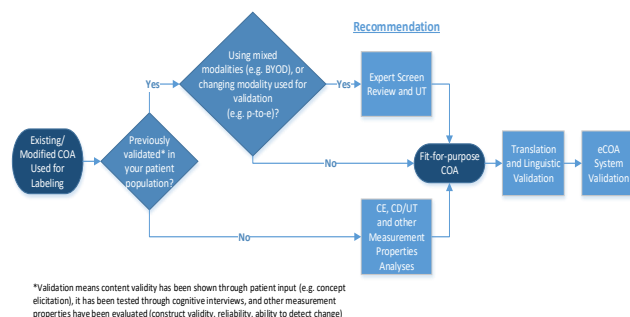


Figure 1: Recommendations for assessment validation of existing or modified COA used to support a labelling claim.

(BYOD=Bring Your Own Device; CD/UT=Cognitive Debriefing and Usability Testing; CE=Concept Elicitation; p-to-e=paper-to-electronic; UT=Usability Testing)

When to use expert screen review

The COA will be used to support a labelling claim, or be used as part of a submission to regulatory agencies. The COA has been previously validated, but it will be faithfully migrated to a different modality (e.g. paper-to-electronic) with minor changes, or mixed modalities (e.g. BYOD) will be used and there is existing evidence (e.g. in the literature) using the same response scale types to support the equivalence between modalities. As with

CD/UT, expert screen review should be performed well prior to phase 3 studies (Figure 1).

MIXED DEVICES

Bring your own device (BYOD)

There is growing interest in allowing patients to use their own devices to collect eCOA data in clinical trials, called BYOD. In a BYOD model, it is important to allow for patients who may not wish to use their own device, as well as those who might lose or damage their personal devices during a study, the means to use a provisioned (study provided) device as a backup. Experience has shown that relying on all patients to comply with a BYOD requirement is rarely, if ever, practical. As such, sponsors may want to consider having a mechanism to provide some number of provisioned devices in a BYOD study as backup. The FDA ‘discussion document for patient-focused drug development public workshop on guidance 3: select, develop or modify fit-for-purpose clinical outcome assessments’ acknowledges the increasing interest in BYOD but still recommends using a single platform throughout a clinical trial to reduce variability.² If a sponsor decides to proceed with BYOD, they should provide a detailed plan to the FDA to review and comment on to ensure that the instrument will function as intended across devices. In the scientific literature it has been recommended that just as expert screen review can apply to paper-to-electronic migration, it can also apply to electronic-to-electronic migrations, such as BYOD.⁹ Therefore, it is recommended that expert screen review and usability testing be performed on the smallest permitted screen size/resolution, and it is also recommended to include a range of devices, including smallest and largest permitted sizes, and both android and iOS devices.^{9,15}

Technical and international customs restrictions

Technical restrictions such as obsolescence of electronic devices/shortage of inventory, or international customs laws preventing device import to certain countries may require eCOA devices to change over the lifetime of a clinical program or be mixed within a clinical trial. In this section, types of technical and international custom restrictions are addressed as well as recommendations to mitigate the risk.

Obsolescence of electronic devices

With improved product development, new technologies, and faster production cycles, manufacturers have the ability to bring new products to the market quickly and force previous products to obsolescence. Experts say that the obsolescence problem is not going to get any better, and that approximately ‘3% of the worldwide electronic components becomes obsolete every month.’¹⁶ Electronic device obsolescence poses a risk to the quality of data collected from clinical trials and may require devices to

be replaced. Hence, determining the equivalence of all devices used in a clinical trial or program becomes a necessity in order to mitigate the risks. Although there are several mechanisms of obsolescence, here are two related to electronic devices:¹⁷

Physical obsolescence: Characterized by products being intentionally designed to last for a certain period of time, allowing limited or restricted repairs, or by products looking old compared to new ones recently launched.¹⁷ Consumers look for the best overall-functioning device; therefore, investing in efficient, innovative, and attractive devices may boost manufacturers' competitiveness while it eliminates obsolete, and sometimes costly, ones. This action also limits consumers' choices and forces them to adapt or switch to new technologies to obtain more benefits.¹⁸

Technical obsolescence: Characterized by introducing new and upgraded products to the market to replace existing ones. In this case, products that are available are still functioning and not in need of being replaced. This type of replacement is considered voluntary as it is based on consumers' demand for satisfaction and manufacturers' demand to satisfy consumers.¹⁷

Shortage of inventory

Manufacturers may decide to halt production of certain devices for many reasons such as 'economics, supply outweighing demand, and consumers' non-satisfaction with the product' among others.¹⁷ In device-provisioning companies, these actions are often out of their control. A shortage of devices is inevitable given that companies cannot order and stock sufficient devices to supply future demand for the duration of a clinical trial that can last for years. In addition, implementation of obsolete devices in a clinical trial may result in a shortage of inventory, therefore it is recommended that dedicated devices have a compatible and equivalent mode of electronic data collection in the event of potential shortages of obsolete devices.

International customs restrictions

International customs laws restrict entry of certain electronic devices into certain countries. These laws require manufacturers or vendors to obtain a regulatory certificate in order to gain country access. Vendors must test their devices against each country's regulatory standards and regulatory agencies use these test reports to prove conformance and grant device certification. Not all devices used for eCOA have obtained regulatory certificates in all countries. However, the eCOA vendor should select devices that are able to enter the majority of countries. In addition, clinical trial sponsors and CROs have the possibility of using clinical trial exemptions to import devices into some countries. This is a time-consuming task and sometimes does not resolve the problem. Nonetheless, these restrictions should not

prevent clinical trial sponsors from testing drugs worldwide or prevent sponsors from electronically collecting data. Instead, alternative devices that have passed countries' regulatory standards, are certified, and found to be equivalent to all other devices should be used in the trial or program. Each eCOA vendor must make sure devices are compliant with international customs restrictions and provide regulatory guidance for each country and each type of device.

Recommendations when using mixed devices within a clinical trial or program

Use of obsolete technology may result in higher failure rates, difficult repairs, product recalls, and increased costs, necessitating that devices be replaced over time.¹⁸ In addition, shortage of inventory or international customs restrictions may require devices to be mixed within a clinical trial or program. Proactive actions such as replacing devices used in a clinical trial and certifying the devices as equivalent may help maintain safety in electronic data collecting modes. If any of the above restrictions cause devices to be mixed within a trial or program, sponsors should perform mode-to-mode equivalence testing as discussed previously (e.g. expert screen review coupled with usability testing) on all devices that will be used for the first time in a mixed-mode trial to keep studies at their safest levels of data collecting.

CONCLUSION

The use of eCOA systems to support a labelling claim should be 1) Shown to be 'fit-for-purpose' by validly and reliably measuring the concepts of interest that are clinically relevant and important to the target patient population, 2) Shown to be equivalent across modes and devices used, and 3) Shown to validly collect data as intended with performance stability over the life of the clinical trial. Development of a new COA that will be used to support labelling should follow the recommendations from the FDA 'Guidance for industry patient reported outcome measures: Use in medical product development to support labeling.'¹ For previously validated COAs that will undergo minor modifications (e.g. paper to electronic migration or electronic to electronic migration such as BYOD or due to technical/customs restrictions), recent scientific literature has recommended performing expert screen review coupled with usability testing to evaluate equivalence. However, it is important to note that expert screen review alone has not been officially endorsed in any regulatory guidance documents. Therefore, it is recommended that sponsors discuss plans for using eCOA with regulatory agencies early in their protocol development. Validation of eCOA both at the assessment level as well as at the system level ensures collection of accurate and high-quality data.

Funding: ERT

Conflict of interest: The authors are employees of ERT. The authors report no other conflicts of interest in this work

Ethical approval: Not required

REFERENCES

1. US food and drug administration. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims. 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); 2009. Available at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims>. Accessed on 28 June 2020.
2. US Food and drug administration. Discussion document for patient-focused drug development public workshop on guidance 3: select, develop or modify fit-for-purpose clinical outcome assessments. 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); 2018. Available at: <https://www.fda.gov/media/116277/download>. Accessed on 28 June 2020.
3. Zbrozek A, Hebert J, Gogates G, Thorell R, Dell C, Molsen E et al. Validation of electronic systems to collect patient-reported outcome (PRO) data recommendations for clinical trial teams: Report of the ISPOR ePRO systems validation good research practices task force. *Value Heal*. 2013;16(4):480-9.
4. US food and drug administration. FDA patient-focused drug development guidance series for enhancing the incorporation of the patient's voice in medical product development and regulatory decision making. 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); 2019. Available at: <https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical>. Accessed on 28 June 2020.
5. European medicines agency (EMA). Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products. 2005. Available at: <https://www.ema.europa.eu/en/regulatory-guidance-use-health-related-quality-life-hrql-measures-evaluation-medicinal-products>. Accessed on 28 June 2020.
6. European Medicines Agency (EMA). Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man the use of patient-reported outcome (PRO) measures in oncology studies. 2016. Available at: <https://www.ema.europa.eu/en/appendix-2-guideline-evaluation-anticancer-medicinal-products-man-use-patient-reported-outcome-pro>. Accessed on 28 June 2020.
7. Souza AC de, Alexandre NMC, Guirardello E de B. Psychometric properties in instruments evaluation of reliability and validity. *Epidemiol e Serv saude Rev do Sist Unico Saude do Bras*. 2017;26(3):649-59.
8. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. *Value Heal*. 2009;12(4):419-29.
9. Muehlhausen W, Byrom B, Skerritt B, McCarthy M, McDowell B, Sohn J. Standards for instrument migration when implementing paper patient-reported outcome instruments electronically: recommendations from a qualitative synthesis of cognitive interview and usability studies. *Value heal*. 2018;21(1):41-8.
10. Byrom B, Gwaltney C, Slagle A, Gnanasakthy A, Muehlhausen W. Measurement Equivalence of Patient-Reported Outcome Measures Migrated to Electronic Formats: A Review of Evidence and Recommendations for Clinical Trials and Bring Your Own Device. *Ther Innov Regul Ici*. 2019;53(4):426-30.
11. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value heal*. 2005;8(2):94-104.
12. US Food and Drug Administration. Part 11, electronic records; electronic signatures-Scope and application. 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), Center for Food Safety and A; 2003. Available at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/part-11-electronic-records-electronic-signatures-scope-and-application>. Accessed on 28 June 2020.
13. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value Heal*. 2008;11(2):322-33.
14. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N et al. Equivalence of electronic and paper administration of patient-reported outcome measures: A systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health qual life outcomes*. 2015;13(1):167
15. Gwaltney C, Coons SJ, O'Donohoe P, O'Gorman H, Denomey M, Howry C et al. Bring your own device (BYOD): The future of field-based patient-reported outcome data collection in clinical trials. *Ther innov regul Sci*. 2015;49(6):783-91.
16. Sandborn P. Design for obsolescence risk

management. In: *Procedia CIRP*. Elsevier BV. 2013;11:15-22.

17. Aladeojebi TK. Planned Obsolescence. *Int J Sci Eng Res*. 2013;4(6):1504-8.
18. Amankwah-Amoah J. Integrated vs. Add-On: A multidimensional conceptualisation of technology obsolescence. SSRN. Published online 2017.

<https://ssrn.com/abstract=2876252>. Accessed on 23 March 2020.

Cite this article as: Gary ST, Otero AV, Faulkner KG, Dias NR. Validation and equivalency of electronic clinical outcomes assessment systems. *Int J Clin Trials* 2020;7(4):271-7.