VALIDATE AN APP: HOW TO DESIGN YOUR STUDY AND GET PUBLISHED

Dr Orrin Franko MD[1,2]

[1]Lead App Editor, Journal of Mobile Technology in Medicine, [2]Dept of Orthopaedic Surgery, University of California, USA.
Corresponding Author: ofranko@ucsd.edu

The last two years have demonstrated an exponential growth in the use of smartphones and tablets by medical professionals, a trend that has led to medical apps developed specifically for patients and physicians.[1-7] Not surprisingly, because most app developers are unverified sources of medical information, recent publications have emphasized the importance of peer-review validation.[7-10] In addition to safety concerns, the validation of mobile apps in the health care setting provides an opportunity for younger physicians, often medical students and residents, to contribute to the medical community by demonstrating the efficacy and validity of these new technologies. However, many trainees and practicing physicians are unfamiliar with scientific validation methodology. This editorial outlines a structure that can be used to assist with the design, execution, and publication of a validation study for mobile technology.

Validation refers to proving a tool's ability to report the absolute "truth" as much as it can be measured. Various forms of validity exist that, when combined, allow a tool to be considered "valid" by the medical community. To clarify various forms of validation, I will share examples from the current literature, which can serve as guides for providers interested in designing a study of their own.

Types of Validation

Before embarking on a validation study, one must possess a clear understanding of the gold standard against which your new tool or app will be validated. A literature search should reveal the existing standard. If not easily identified, consult with a colleague or professional in your field to guide you.

Criterion Validity

Once the gold standard has been selected, criterion validity is the method used to demonstrate a direct correlation between the new tool and existing standard using an appropriate statistical test, such as a Pearson correlation. Consider one study that validated the use an Android smartphone for gait analysis and confirmed criterion-validity by evaluating the correlation between the gait parameters obtained by the smartphone and the gold standard, a tri-axial accelerometer. The statistical test they used was Spearman's correlation coefficient r.[11] Their results demonstrated a correlation between the smartphone and the goniometer ranging from 0.82-0.99, suggesting a strong relationship and thereby confirming their criterion validity.[11]

Construct Validity

Construct validity is another form of validity, and refers to the systematic change in results when the input variable is under varying conditions.[12] More specifically, it answers the question, 'does the new tool do what it is supposed to do?' In contrast to a comparison against an existing standard, construct validity aims to demonstrate an appropriate response against a real-world measure. For example, a recent study validating a virtual reality simulator for robotic surgical skills demonstrated construct validity by correlating outcomes using the device with each participant's level of robotic surgery experience.[13] By demonstrating that experience correlated with their simulator skills, construct validity was established. Importantly, while a tool may not meet criterion validity (it may not measure the desired outcome very accurately), it could still meet construct validity (fulfilling the predicted effect). For these types of comparisons, analysis of variance (ANOVA) is often used to reveal the effect of a single variable when multiple variables are being tested.

Intra-observer Reliability

A third statistical characteristic is intra-observer reliability, also known astest-retest reliability, which reflects a highly-reproducible outcome when tested under constant conditions by the same observer. From a clinical perspective, this implies that results should remain the same when testing conditions are unchanged. For example, one study examined intra-observer reliability when utilizing a smartphone to assess shoulder range of motion by testing 41 subjects twice, with a 30-minute interval between tests.[14] The intraclass correlation coefficient (ICC) was the statistical test used to compare results at the two testing points and revealed a high degree of correlation for each observer, thus confirming intra-observer reliability.

Inter-observer Reliability

Similarly, inter-observer reliability, also known as inter-rater reliability, reflects the accuracy and precision of a tool when used by various care providers. For example, a new device would not be particularly useful if only the developer could use it properly. Thus, it is important to prove that a tool can be equally effective with a basic level of training for different providers. Using the same example as above, the authors also examined inter-observer reliability by testing 3 different providers using the device on the same group of patients. Once again, ICC was used to compare the results and revealed a strong correlation.[14]

Content Analysis

In addition to the statistical validation techniques described above, the content of information provided in apps can be verified by performing a content analysis. In this way, the data within an app is compared to a reliable source, such as a gold standard textbook or guideline. One study performed a content analysis on 47 apps that were advertised to assist with smoking cessation and were evaluated based on their adherence to the U.S. Public Health Service's 2008 Clinical Practice Guideline for Treating Tobacco Use and Dependence. From their analysis, they were able to rate the apps with respect to adherence to the published guidelines.[15]

Study Design and Analysis

Once the validation tests and techniques are understood, you can determine the study methodology. Importantly: are patients required for you to validate your tool? If so, an ethics committee or institutional review board (IRB) must approve your protocol. Keep in mind, however, that many institutional review boards offer an accelerated application for projects that present little risk to subjects. The IRB process includes thinking about patient recruitment, which is often the time-limiting step for a study. Ask colleagues for help, advice, and ideas if you anticipate this will be a challenge. Lastly, focus on how you will collect your data. What data will you collect? How will it be collected and how will the results be stored? Who is collecting the data and who is analyzing it? Will the process be blinded? A great amount of time can be saved by carefully outlining the research plan.

After the plan has been outlined and an IRB application has been approved, data collection should proceed smoothly and efficiently. A trial data collection period will help identify any potential methodological limitations planned in the study. In other words: do not expect your first trials to produce usable data; your measurement techniques are likely to change significantly within the first 5-10% of data collection.

Once collected, data must be analyzed. Statistical analysis intimidates many researchers who are unfamiliar with these tests. If this is true, ask a friend or colleague to help. As outlined above, the general principles for validating new tools do not require particularly difficult statistical tests and can usually be completed after only 1 or 2 meetings with a knowledgeable colleague.

Manuscript Preparation and Submission

The final step is manuscript preparation. They key to writing a compelling and interesting manuscript is allowing the data to drive the study's conclusions in the context of the aims and hypothesis that were set out from the start. Avoid the temptation of trying to fit the data to your conclusion. Rather, recognize that all scientists embark on studies to either confirm, or refute a theory, but it is the unpredictable nature of science that appeals to so many researchers. Examine the data with an open mind, share it with colleagues, and let your results guide your conclusions without bias.

The conventional scientific paper format for nearly all journals is: Introduction/Background, Aims/Hypothesis, Methods, Results, Discussion, and Conclusions. However, the order of preparing each section should not necessarily follow the order of formatting. Rather, a manuscript should typically be written in the following order: figures, results, methods, and discussion, with the introduction written last. Following this sequence most closely represents the intellectual progression of an experiment and can help organize the author's thoughts. The data (figures

and results) are reported, the methods are confirmed, and the implications are discussed and supported. Only *after* a study is completed can an appropriate introduction be written. This step can potentially save hours of revision time.

After reviewing and improving the manuscript, submission to an appropriate journal should not be delayed. Selecting the proper journal also requires care, and important factors to consider include a journal's primary focus, the breadth of readership, publication format (online or print), indexing databases, impact factor, publication costs, copyright ownership, and duration of peer review.

Conclusion

In addition to the many examples described above, there are a number of other good validation studies that can help guide the design of future studies. I would encourage interested readers to read more about a smartphone heart rate acquisition application,[16] an evidence-based application for treating cervical spine trauma,[17] validation of heart rate extraction using an iPhone accelerometer,[18] validation of a Timed Up and Go test,[19] using smartphones to measure Cobb angles in scoliosis,[20-21] and improving total hip arthroplasty component placement with a smartphone.[22]

While the editors of the jMTM take pride in our rapid manuscript review process (often less than 1 month), most journals will take anywhere from 3-6 months (or longer) for the first round of peer-review. As the lead app editor of jMTM, I look forward to reviewing your studies about mobile applications in healthcare.

References

1. Azark R. Smartphone apps for your practice. CDS Rev 2011;104:12-13.

2. Bhansali R, Armstrong J. Smartphone applications for pediatric anesthesia. Paediatr Anaesth 2012;22:400-404.

3. Franko OI. Smartphone apps for orthopaedic surgeons. Clin Orthop Relat Res 2011;469:2042-2048.

4. Franko OI, Bhola S. iPad apps for orthopedic surgeons. Orthopedics 2011;34:978-981.

5. Oehler RL, Smith K, Toney JF. Infectious diseases resources for the iPhone. Clin Infect Dis 2010;50:1268-1274.

6. Rosser BA, Eccleston C. Smartphone applications for pain management. J Telemed Telecare 2011;17:308-312.

7. Franko OI, Tirrell TF. Smartphone App Use Among Medical Providers in ACGME Training Programs. J Med Syst 2011.

8. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011;10:24.

9. Hamilton AD, Brady RR. Medical Professional Involvement in Smartphone Apps in Dermatology. Br J Dermatol 2012.

10. Kabachinski J. Mobile medical apps changing healthcare technology. Biomed Instrum Technol 2011;45:482-486.

11. Nishiguchi S, Yamada M, Nagai K, Mori S, Kajiwara Y, Sonoda T, Yoshimura K, Yoshitomi H, Ito H, Okamoto K, Ito T, Muto S, Ishihara T, Aoyama T. Reliability and Validity of Gait Analysis by Android-Based Smartphone. Telemed J E Health 2012.

12. Smith MV, Klein SE, Clohisy JC, Baca GR, Brophy RH, Wright RW. Lower extremity-specific measures of disability and outcomes in orthopaedic surgery. J Bone Joint Surg Am 2012;94:468-477.

13. Perrenot C, Perez M, Tran N, Jehl JP, Felblinger J, Bresler L, Hubert J. The virtual reality simulator dV-Trainer((R)) is a valid assessment tool for robotic surgical skills. Surg Endosc 2012.

14. Shin SH, Ro DH, Lee OS, Oh JH, Kim SH. Within-day reliability of shoulder range of motion measurement with a smartphone. Man Ther 2012.

15. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011;40:279-285.

16. Gregoski MJ, Mueller M, Vertegel A, Shaporev A, Jackson BB, Frenzel RM, Sprehn SM, Treiber FA. Development and validation of a smartphone heart rate acquisition application for health promotion and wellness telehealth applications. Int J Telemed Appl 2012;2012:696324.

17. Kubben PL, van Santbrink H, Cornips EM, Vaccaro AR, Dvorak MF, van Rhijn LW, Scherpbier AJ, Hoogland H. An evidence-based mobile decision support system for subaxial cervical spine injury treatment. Surg Neurol Int 2011;2:32.

18. Kwon S, Lee J, Chung GS, Park KS. Validation of heart rate extraction through an iPhone

accelerometer. Conf Proc IEEE Eng Med Biol Soc 2011;2011:5260-5263.

19.   Mellone S, Tacconi C, Chiari L. Validity of a Smartphone-based instrumented Timed Up and Go. Gait Posture 2012.

20.   Qiao J, Liu Z, Xu L, Wu T, Zheng X, Zhu Z, Zhu F, Qian B, Qiu Y. Reliability Analysis of a Smartphone-aided Measurement Method for the Cobb Angle of Scoliosis. J Spinal Disord Tech 2012.

21.   Shaw M, Adam CJ, Izatt MT, Licina P, Askin GN. Use of the iPhone for Cobb angle measurement in scoliosis. Eur Spine J 2011.

22.   Peters FM, Greeff R, Goldstein N, Frey CT. Improving Acetabular Cup Orientation in Total Hip Arthroplasty by Using Smartphone Technology. J Arthroplasty 2012.