FDA U.S. FOOD & DRUG ADMINISTRATION
CENTER FOR DEVICES & RADIOLOGICAL HEALTH

Challenges in Validating Mobile Medical Applications

Arkendra De, FDA-CDRH

Saryet Kucukemiroglu, FDA-CDRH

2020 ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop

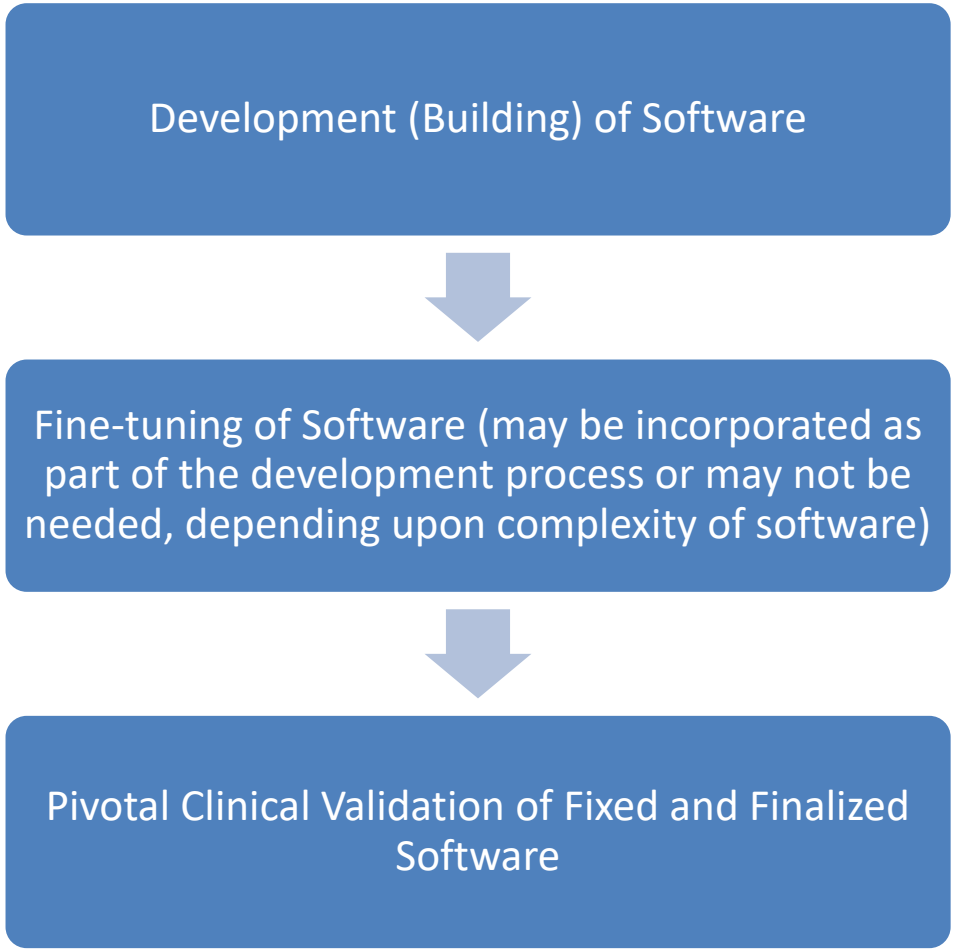
Outline

- What are Mobile Medical Applications?
- Two Key Considerations for Clinically Validating Mobile Medical Applications
- Challenges for Clinical Validation
- Summary

What are Mobile Medical Applications (MMAs)?

- Software that is meant to be installed directly on mobile devices (e.g. smartphones, smartwatch, tablet) for the purpose of interfacing with other hardware (e.g. internal hardware on the phone, hardware external to the phone) that can be used for measuring or providing medical information.
- The medical information provided by the software can vary depending upon the type of device, where the software is installed, and the functionality of the software.
- Examples: balance application that provides scores for different balance exercises; cardiovascular application for detecting atrial fibrillation (heart condition); mobile application for identifying the owner of dental dentures in the situation where they are lost
- More examples can be found in [1], [2].

Stages for Mobile Medical Application Construction



The bottom box in the above figure will be the focus and scope of this presentation.

Why the focus on pivotal clinical studies?

- Mistakes could be made in the internal validation of software (i.e. assessment of software performance using the data collected for developing, training, and/or fine-tuning the software) [3], [4], [5].

- Estimates of performance obtained during development may not be generalizable if the data collected for developing software are not representative of the expected use of the software [6], [7].

- Some have said that internal validation has other limitations [6], [7].

- For these reasons, it is important to evaluate the external validity of the software (external validity: “ability to maintain accuracy when applied to patients and settings different from those on which the models were developed” [6]), which is the aim of pivotal clinical validation studies.

Two Key Aspects for Pivotal Clinical Validation Studies

1) Pivotal clinical accuracy study

- Assessment of the concordance between the MMA and clinical reference standard for a particular target condition of interest when the study is designed and conducted according to a particular context of use and setting
- Clinical reference standard is best available method for ascertaining the target condition of interest [8] and is determined by the clinical community and is expected to differ for different target conditions of interest.
- Using the framework in slide 4, MMA is fixed and finalized before commencing the pivotal clinical accuracy study to avoid bias [8], [9]. To avoid bias, the pivotal clinical study uses individuals completely separate from those used for developing the MMA [8], [9].
- Usually for clinical accuracy, paired design is used, where measurements are recorded from the reference standard and MMA for each subject.

Two Key Aspects for Pivotal Clinical Validation Studies

1) Pivotal Clinical Accuracy Study

- When the MMA and clinical reference standard (CRS) can be considered as binary (e.g. positive for the target condition of interest, negative for the target condition of interest), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) point estimates are computed, along with confidence intervals [10].

Sensitivity = Pr(MMA+ | CRS+)

Specificity = Pr(MMA- | CRS-)

PPV = Pr(CRS+ | MMA+)

NPV = Pr(CRS- | MMA-)

Two Key Aspects for Pivotal Clinical Validation Studies

- Typically for MMAs with a continuous output, Deming regression analysis [11] and Bland-Altman plot [12] are performed.

Note: In recognition of the importance of the clinical reference standard for defining clinical accuracy, if a comparator used for a study is not the clinical reference standard, then the study is called as a clinical agreement study [8]. Similar study design could be used as described on the previous slide, but the nomenclature for statistical analyses changes to recognize the change in comparator [10].

Two Key Aspects for Pivotal Clinical Validation Studies

2) Measurement Validation

- Characterization of various aspects associated with a medical test's ability to measure what the medical test is designed to measure
- Multiple assessments and studies for medical tests are considered to be part of measurement validation [9], including precision.
- Precision for all medical tests is "the closeness of agreement between replicate measurements on the same object (e.g., sample, subject) under specified testing conditions" [9], [13].
- Multiple variables (factors) usually are included for a precision study (e.g. operator, medical test making measurements, subject, etc.).

Two Key Aspects for Pivotal Clinical Validation Studies

- Goal of statistical analyses for precision studies is to separate out and estimate variability due to measurements made by medical test from the variability due to other factors/variables (e.g. operator, subject being tested)

- To accomplish this goal, typically use random effects models for statistical analyses of precision studies.

Challenges with Pivotal Clinical Accuracy Studies for MMAs

- Study enrolling a non-representative sample of individuals for assessment of clinical accuracy

  Example: inflating the percentage of individuals with target condition of interest for a pivotal clinical study compared to what would be expected in the target population at large (called as an enriched study)

  Example: enrolling only a particular subgroup (e.g. males) for a pivotal clinical study

Challenges with Pivotal Clinical Accuracy Studies for MMAs

Remarks about the use of a non-representative sample:

a. Non-representative sample could introduce spectrum bias for the study design.

b. Spectrum bias refers to the phenomenon where the spectrum of target condition of interest for the pivotal clinical study may not be representative of the target population at large [14-16].

c. Spectrum bias is known to introduce bias for sensitivity, specificity, PPV, and NPV [14-16].

Challenges with Pivotal Clinical Accuracy Studies for MMAs

- Using case-control study design for the pivotal clinical accuracy study

  a. Raises concern about spectrum bias [14-16].

  b. Could introduce bias for estimates of clinical performance [14-19]

  c. Other concerns with the use of case-control study designs for pivotal clinical accuracy studies are in [17- 19]

Challenges with Pivotal Clinical Accuracy Studies for MMAs

- Use of stored data from registries or databases for clinically validating MMAs, where there is some type of comparator and MMA result on the same subjects collected from the past (here, subjects for development of MMA are completely separate and independent from the stored data proposed for the pivotal clinical accuracy study)

a. If the comparator used varies across subjects, then this could introduce bias in the comparison between the MMA and comparator. This situation also introduces difficulty in interpreting results since the comparator is inconsistent.

b. Version of MMA used for the stored data could differ from the current version of the MMA, which could introduce difficulties for generalizing results to the target population at large for the current version of the MMA

c. Other concerns with the use of stored data are given in [9].

Challenges with Pivotal Clinical Accuracy Studies for MMAs

- Going back to the same stored dataset for assessing the external validity of clinical accuracy for new versions or iterations of MMA (here, subjects used for development of MMA for all versions of MMA are all completely separate and independent from the stored dataset described here)
    a. Use of stored data for assessing external validity may not be independent of development since knowledge of the validation set (i.e. stored dataset described in the above bullet point), indirectly or directly, may influence development; could introduce bias

    b. Using the same stored dataset for validating new versions of the MMA may not reflect the characteristics of the current target population and/or current clinical reference standard since these characteristics are expected to change over time.

    c. Other concerns with the use of stored data are given in [9].

Challenges with Pivotal Clinical Accuracy Studies for MMAs

- Random splitting of master data set into training and validation

    a. Inefficient use of data since not all data are used for validating [20], [21].

    b. Creates more similarity between training and validation sets than what would be expected if an independent clinical study were performed for assessing the external validity of the MMA [20], [21]. This is because randomization balances observed and unobserved variables, which can introduce difficulty in generalizing results [20], [21].

Challenges with Pivotal Clinical Accuracy Studies for MMAs

- Use of post-hoc analyses (i.e. analyses performed that were motivated or inspired only after examining the contents of the data) of the pivotal clinical accuracy study (e.g. subgroup analyses) for supporting clinical performance claims of the MMA
  a. Jeopardizes the integrity of results and pivotal clinical study [8]
  b. Could potentially inflate statistical type I error [8]
  c. Raises questions about generalizability of results to the target population at large due to potentially overoptimistic estimates of performance [8]

Challenges with Pivotal Clinical Accuracy Studies for MMAs

- Tailoring/modifying MMA using data from pivotal clinical accuracy study
  a. Jeopardizes the integrity of results and pivotal clinical study [8]
  b. Could potentially inflate statistical type I error [8]
  c. Raises questions about generalizability of results to the target population at large due to potentially overoptimistic estimates of performance [8]

Challenges with Clinical Precision Studies for MMAs

- Separating sources of variability so that variability of measurements due to MMA is assessed

  Example: For assessing balance, MMA is designed so that mobile device is attached or held by the individual whose balance is being tested. Since the operator of the MMA is the one being tested, it is difficult to separate subject variability from the variability due to operator and the variability due to the measurements made by the MMA for the statistical analyses. All three of these factors are conflated together.

- Most MMAs are marketed directly to consumers (those without formal healthcare training). Differences in the manner of operation of the MMA due to the user and/or due to the type of mobile device used could introduce additional variability for the precision study results and may not be possible to separate from the variability due to the measurements made by the MMA for the statistical analyses.

  Example: hearing app

Summary

- MMAs are a diverse class of products that are increasing in importance
- Validation (i.e. assessing external validity) of MMAs has been and will continue to be an essential component
- Using independent sets of subjects for validating and developing MMAs is not enough – important to think and plan regarding where, how, and what data are collected for validation and also what previous activities were conducted with the proposed validation data

Note

- The challenges discussed in the preceding slides are not meant to be exhaustive or comprehensive. Also, the concerns raised for each of these challenges are not meant to be exhaustive or comprehensive. Rather, the preceding slides are meant to highlight some of the most common challenges and some of the major concerns encountered for the validation (i.e. assessment of external validity) of MMAs.
- Also, the challenges for the clinical accuracy study discussed in the preceding slides also are applicable to other medical tests. The references listed in the next two slides provide details about this.

References

1. U.S. FDA MMA Examples (2020) (https://www.fda.gov/medical-devices/device-software-functions-including-mobile-medical-applications/examples-premarket-submissions-include-mmas-cleared-or-approved-fda)
2. U.S. FDA (2019) Policy for Device Software Functions and Mobile Medical Applications https://www.fda.gov/media/80958/download
3. Qin LX, Huang HC, Begg CB (2016) Journal of Clinical Oncology 34(32):3931-3938. doi: 10.1200/JCO.2016.68.1031
4. Moons, KGM et al (2012). Heart Published Online: 07 March 2012 doi: 10.1136/heartjnl-2011-301247
5. McShane, LM and Polley, MYC (2013) Clinical Trials 10:653 doi: 10.1177/1740774513499458.
6. Terrin, N et al. (2003). Journal of Clinical Epidemiology 56: 721–729 doi: 10.1016/S0895-4356(03)00120-3
7. Bleeker, SE et al (2003) Journal of Clinical Epidemiology 56 (2003) 826–832 doi: 10.1016/s0895-4356(03)00207-5
8. U.S. FDA (2013) Design Considerations for Pivotal Clinical Investigations for Medical Devices. (https://www.fda.gov/media/87363/download)
9. De, A et al (2013). Journal of Cardiovascular Translational Research 6:449-457 doi:10.1007/s12265-013-9470-3
10. U.S. FDA (2007) Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests. (https://www.fda.gov/media/71147/download)
11. Konings, H (1982) Survey of Immunologic Research 1(4):371-4 doi: 10.1007/BF02918550.
12. Bland JM, Altman DG. (1986). Lancet, i, 307-310 doi: 10.1016/S0140-6736(86)90837-8
13. CLSI. (2014). Evaluation of precision performance of quantitative measurement methods; Approved guideline–third edition. CLSI document EP5-A3. Wayne: Clinical and Laboratory Standards Institute

References

14. Usher-Smith, JA, Sharp, SJ, Griffin, SJ (2016) BMJ 353:i:3139 DOI: 10.1136/bmj.i3139.

15. Moons, K.G.M. and Harrell, F.E. (2003). Academic Radiology, 10, 670-672. doi: 10.1016/S1076-6332(03)80087-9

16. Brenner, H, Gefeller, O. (1997) Statistics in Medicine 16: 981-991 doi: 10.1002/(SICI)1097-0258(19970515)16:9<981::AID-SIM510>3.0.CO;2-N

17. Pepe, MS at al (2012) Clinical Chemistry 58:8 1242–1251 doi: 10.1373/clinchem.2012.186007

18. Rutjes, AWS et al (2005) Clinical Chemistry 51:8, pp. 1335–1341 doi: 10.1373/clinchem.2005.048595

19. Lijmer JG et al (1999) JAMA 282:1061–1066 doi: 10.1001/jama.282.11.1061

20. Vishnuvajjala, RL (2015) ”Issues with Training, Testing and Validation Datasets in the Development of Diagnostics”. Proceedings of 2015 Joint Statistical Meetings, held at Seattle, WA.

21. Moons, KGM et al (2015) Annals of Internal Medicine 162:W1-W73 doi:10.7326/M14-0698.