Mobile Health Applications, in the Absence of an Authentic Regulation, Does the Usability Score Correlate with a Better Medical Reliability?

Mobin Yasini¹, Guillaume Marchand¹,²

¹ DMD santé, Research and development department, 51100, Reims, France
² Department of psychiatry, Rouen University Hospital, 76031 Rouen Cedex, France

Abstract

Health-related mobile applications (apps) have been shown to improve the quality of health and patient care. Their use in clinical and health-related environments is becoming more considerable. The number of health-related apps available for download has considerably increased, while the regulatory position of this new industry is not well known. Despite this lack of regulation, measuring the usability score of these apps is not difficult. We compared two samples of twenty health-related applications each. One of the samples contained the apps with top-rated usability scores, and the other contained the apps with lowest-rated usability scores. We found that a good usability score correlates with a better medical reliability of the app’s content (p<0.005). In the period in which a valid regulation is still lacking, calculation and attribution of usability scores to mobile applications could be used to identify apps with better medical quality. However, the usability score method ought to be rigorous and should not be rounded off with a simple five stars rating (as is the case in the classic app stores).

Keywords:

mHealth, Mobile Applications, Regulation, Smart Phones, Equipment Safety.

Introduction

Mobile technology solutions have an important place in daily life today. In addition to simple communication and text messaging, a variety of applications running on smartphones and tablets could offer easily accessible solutions to almost any possible requirement that may arise in the field of information and communication. Health-related information and communication is not an exception in the exploding penetration of mobile technology phenomenon.

Mobile Health – or mHealth – is the practice of medical or public health supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices [1]. Service care providers, researchers, and national governments are excited at the opportunities mobile health has to offer in terms of improving access to health care, engagement and delivery, and health outcome [1]. Numerous functionalities including data collection [2], patient communication [3], health care delivery [4,5], and patient and professional education [6,7] have been used to support patients or healthcare professionals.

Currently, more than 97,000 mHealth applications are present in the classic app stores, and approximately 1,000 new applications are being published every month [8]. In the next few years, more than 3,000,000 free downloads and 300,000 paid downloads are expected to be made of mHealth applications just in the USA [9]. Being easily accessible and highly available make smart devices (smartphones and tablets) very attractive for consumer and professional areas of application [8]. The smartphone is considered to be the most popular technology among physicians since the stethoscope [9]. Therefore, there are high expectations for growth in the mHealth market. With this growing market and the overwhelming diversity of applications, there is a need to recognize which applications are safe and trustworthy. In the recent literature, potential dangers and reliability of some health-related applications were investigated [10–12]. Various aspects of mobile health applications have been called into question, including medical professional involvement in the development of the applications [13], the accuracy and reliability of the content of the applications used in diagnosis and patient management [14], the potential danger of camera function in mobile devices to judge weather skin lesions are suspicious [15,16], deficiencies in self-management applications in diabetes [17] and asthma [18].

The mobile health application industry is still in its infancy. Existing laws and public regulations for approving health-related applications are only relevant to a rather small number of applications [8]. Happtique, a private certification of health-related applications suspended its operation after a developer discovered that two certified apps contained data insecurity issues [19]. Various reasons, including the number of functionalities, diversity of information and rapid development of health-related applications in the market, make certification difficult to achieve [20].

A usability score indicates the satisfaction and effectiveness of an application. Major platforms for mobile applications (Google Play store & Apple store) provide consumers the ability to rate the applications. Users can rate the applications by giving them a mark out of five stars. In the absence of a comprehensive and rigorous certification of mobile applications for health care, the only easily accessible quantitative measurement is the usability score. The main aim of this study was to evaluate if good usability scores correlates with a better medical safety. We therefore set out to determine if health-related apps with disparate usability scores differ in medical safety aspect.

Methods

French language-based health-related mobile apps in classic app stores (Google play store and Apple iOS app store) were gathered on our website at www.dmdpost.com. A monitoring system based on keywords extracted health-related applications from the classic app stores. App editors who would like to have their applications evaluated by our community could also directly send us their apps. The website currently contains 896 mobile health applications. These applications were divided into two categories according to
their target user: health professionals and the general public. A panel of application users (health professionals and the general public including patients) voluntarily completed usability satisfaction questionnaires for each mobile application. Everyone (health professionals, patients, and healthy individuals) could evaluate the apps addressed to the general public. However, if the target users of an app are health professionals, it would be evaluated solely by its target users. There were three questionnaires according to the evaluator’s profile:

- A questionnaire for patients or healthy individuals to evaluate the apps intended for the general public
- A questionnaire for health professionals to evaluate the apps intended for the general public and to identify whether the health professional is willing to prescribe the app to his/her patients
- A questionnaire for health professionals to evaluate the apps intended for health professionals to verify if the application matches the healthcare professionals needs

There are two types of questions in each satisfaction survey both in a closed-ended format:

- Influence questions: to quantify application usage and reliability of answers (for example, Did you use the application several times a day for more than one week?)
- Perception questions: to assess the quality of the application on its functionality and ergonomic aspects (for example, Do you find the application easy to use?)

When the number of evaluations reaches five (at least four individuals and one health professional), the embedded algorithm assigns a score out of 20, judging the overall quality of the application. If an application is updated, it will pass to the status “Rating in progress” and will be re-evaluated by at least four individuals and one health professional to mark its current version.

We then selected the 20 highest-rated and the 20 lowest-rated applications. These applications constitute the material for this study. Two medical doctors working on mHealth evaluated four basic issues to assess medical safety of these applications. The four issues are:

- Presence of at least one related health professional in the conception or development of the application
- Using reliable and valid bibliographic references to create the application contents
- Clear identification of objectives, target audience, and/or mentioning the possibility (or no possibility) of substituting a medical advice within the application
- Validation of the application in a peer-reviewed scientific paper

These criteria were then validated by a group of academic experts in the relevant field. The best rated and lowest rated applications were listed randomly so that the evaluators did not know the applications’ usability scores during their evaluation process. The evaluators examined independently each application with the basic criteria. Discordances were then discussed to obtain a consensus. The name of each application was searched in Medline and Web of Science databases to identify if the application was the subject of a peer-reviewed published research. We then compared the total scores in the two groups. Descriptive information was produced for each basic issue of medical safety. The results were analyzed using Chi-square test. Significance of the statistic test was set at \( p<0.05 \).

Next, we compared the given app store usability scores of the two groups. In this step, we were interested in finding out if the user evaluations out of five stars could be considered as a reliable criterion to predict the medical reliability of the applications. The user evaluations out of five stars, published in the classic app store platforms within the app description, were analysed and compared in the two groups. We used the Wilcoxon rank-sum test to compare the mean score in the two groups.

### Results

The applications used in this study covered various subjects including allergies, contraception, pregnancy and menstrual cycle, calorie intake, bipolar disorder, multiple sclerosis, medical or therapeutic education, cardiology scores, general follow up, diabetes mellitus, traveler’s health, and hepatogastroentrology recommendations. The target users of these applications were health professionals and the general public including patients. The majority of apps in both groups were intended for the general public. However, the number of apps intended for health professionals was more prevalent in the highest-rated group than the lowest-rated group (40% of the apps in the top-rated group against 5% of the apps in the lowest-rated group).

The number of concordances with our medical safety criteria together with the relevant percentages are illustrated in table 1.

<table>
<thead>
<tr>
<th>Highest rated applications (%)</th>
<th>Lowest rated applications (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of related health professional</td>
<td>14 (70%)</td>
</tr>
<tr>
<td>Reliable references</td>
<td>11 (55%)</td>
</tr>
<tr>
<td>Identification of objectives, target audience, possibility of replacing a medical advice</td>
<td>16 (80%)</td>
</tr>
<tr>
<td>Validation of the application in a peer reviewed scientific paper</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Total</td>
<td>41 (68.3%)</td>
</tr>
</tbody>
</table>

Presence of related health professionals in the conception or development of the application was observed more in the top-rated applications than the lowest-rated applications. The criterion satisfied in the fewest number of applications in both highest-rated and lowest-rated groups was mentioning whether reliable or valid bibliographic references were used to create the application contents (11 in top-rated applications and 0 in lowest-rated applications). The third criteria fulfillment (identification of objectives) was observed equally in both top-rated and lowest-rated applications (16 out of 20 for each group). In the top-rated group, we found a total of 41 times (68.3%) satisfaction to our medical safety criteria versus 24 times (40%) in lowest-rated applications. None of the studied applications were the subject of a peer-reviewed paper to assess its effectiveness. We therefore decided to remove this criterion from our calculations.
According to our criteria, the medical reliability was significantly higher in the highest-rated applications compared to the lowest-rated applications \( (p<0.005) \).

In comparing an application’s usability score to the classic app store rating, we found that the highest-rating group have a mean score of 3.87 (standard deviation = 0.67). In the highest-rated group, four applications were removed due to an insufficient number user ratings for the application. The range of marks was between 3.5 and 5. In the lowest-rated group, six applications were removed due to insufficient user ratings from the classic app store. The mean score for the 14 remaining was 3.78 (standard deviation = 1.03). The range of ratings in this group was between 2 and 5. We did not find a statistically significant difference between these two means ratings \( (p<0.9) \) in our samples. We therefore did not compare the store provided usability scores with our medical criteria.

**Discussion**

Mobile applications comprise a flourishing industry today. Health-related applications provide users with an unprecedented opportunity to achieve a better quality of health as well as a better quality of healthcare. However, when it concerns health and healthcare goals, one must consider the potential risks and dangers. The ubiquity of mobile applications – including health-related apps – brings additional risks to less-experienced users who might download applications with dubious medical information and advice.

Although some voluntary app certification schemes like “Happtique Health App Certification Program” exist, a worldwide comprehensive and rigorous certification or regulation in all possible aspects concerning the mobile health applications including privacy, operability, security, ethical issues, medical content reliability, etc., is still lacking. The traditional methods of evaluation are not suitable for the fast paced nature of technology. However, calculating a usability score is more common and has already been used in the published investigations. In this study we present an analysis of 40 health-related applications to assess the feasibility of judging the medical reliability of mobile health-related applications by their usability score. In particular, we developed a usability evaluation method that was more rigorous than the probabilistic system out of five stars available on the classic store platforms. We defined four basic criteria to assess the medical content of health-related applications. We found that these criteria are significantly more fulfilled in the applications with a better usability score than the applications with lower usability score.

Other studies that evaluated the usability of health-related applications admitted that good usability scores in applications are related with better results in functionality and high compliance of consumers. This corroborates the results of our study, which report that good usability scores correlate with better medical safety.

We then compared the store platform provided usability score in our samples. Our findings showed that the store-provided usability score was not significantly different in the two groups. A simple consumer rating out of five star is therefore not a good benchmark for evaluation of the usability of health-related apps.

This study had several limitations, which we tried to mitigate. We blinded evaluators of applications’ usability scores to reduce bias during their evaluation of medical criteria.

The rate of being approved by a peer-reviewed publication was poor in both highest-rated and lowest-rated groups of apps. This is in line with another published study in which among 283 pain-related apps, none of them had been scientifically validated or proven to be effective. However, our sample applications were only in French language. This can reduce the chance of being the subject of a peer-reviewed publication, because most of these publications are in English language and assess rather the English language applications.

The use of reliable references in the app development was not sufficiently mentioned for any of the apps, even in the highest-rated applications. This defective feature could be somewhat compensated by having a licensed health care professional implicated in the development of the application. Furthermore, valid and reliable sources might have been used in some applications without being mentioned.

Presence of related health professionals in the development of the application was found to be the most prevalent source of support to health-related apps. This finding corroborates with another study who reports the implication of health care professionals in the app development followed by being recommended by a patient association as the most important source of support to the apps. An analysis of colorectal disease smartphone apps in another study showed that only 29% of the apps had had customer satisfaction ratings and 32% had named medical professional involvement in their development or content. This finding is in line with our results in lowest-rated health-related apps with 40% of medical professional involvement.

The present user evaluation system provided by the store platforms seems not to be a good reference to conclude if a health-related app is medically safe. This study demonstrated that a more rigorous system to measure usability score of the health-related apps could correlate with a better medical reliability. It is not yet clear how the assessment for health-related apps should best proceed. In this situation, although the correlation with usability score is not an absolute and precise solution, it can be used in the absence of a comprehensive regulation.

The proportion of apps intended for health professionals was found to be more prevalent in the highest-rated group. It seems that app editors or developers pay more attention to the content when the target consumers are health-literate professionals. However the majority of health-related applications are intended for the general public (as is also evident in our samples) who may not have the same health literacy and numeracy of health professionals. Educational programs for health-related app users could not only empower patients and healthy individuals to have a better insight of the apps that they are using, but also arouse app editors to develop better quality and reliable apps.

Perhaps the most important limitation of this study is that we only looked at French language applications. This could have limited the yield of the results. However, some of these apps were present in the stores in a multilingual format. Small sample size is another limitation. We selected only 40 of the possibilities, not only because it was convenient, but also because it was not feasible to incorporate the apps. In this study the apps were taken from all the diverse categories of health-related apps that could be found. As the different health fields have different characteristics, further studies analysing apps by dividing them according to their purposes or functionalities would be of considerable interest.

Even in the top-rated group, a great percentage of applications (more than 20%) lacked at least one of the medical criteria. This implies the need for a comprehensive multi axial program to assess the health-related apps in all aspects. Medical trustworthiness of health-related apps is of utmost importance, however high-quality, evidence-based content
alone is of limited value, if presented in a way that does not adequately match and address the usability, accessibility, readability (reading with understanding) and health literacy needs of targeted audiences [21].

**Conclusion**

Mobile health technology seems to be a logical, acceptable, and efficient way to improve the quality of health and health care services. It can make healthcare more accessible and affordable for all [9]. It has been demonstrated that smartphone-based applications improve the efficiency of healthcare delivery and make healthcare more effective [27]. To achieve this objective, it is clear that the content of applications used in a medical or health context are required to be trustworthy. Mobile health applications still comprise a new industry. Therefore, it takes some time to achieve a valid comprehensive assessment procedure to proceed the app review. Meanwhile, calculation of usability scores that are more common among mobile applications could be used to distinguish apps with better medical quality. However, the usability score method ought to be rigorous and should not be rounded off with a simple five stars rating. Further research with more important samples would help to confirm the results of this study and pave the way to find trustworthy methodological issues to assure the quality of mobile health.

**Acknowledgments**

The authors gratefully acknowledge the contribution of Mr. Gregory Duval and Mr. Bastien Burlon for their recommendations. There is no financial or personal relationship with organizations that could potentially be perceived as influencing this research.

**References**


Address for correspondence
Mobin YASINI, DMD Santé, Research and development department, 38 rue Cérès, 51100, Reims, France.
Tel : +33 9 73 19 81 19; Email : mobin.yasini@dmd-sante.com