Original Paper

App Usage Factor: A Simple Metric to Compare the Population Impact of Mobile Medical Apps

Thomas Lorchan Lewis[1], BSc (Hons), MB ChB; Jeremy C Wyatt[2], BA, MBBS, DM, FRCP, FACMI

[1]Kingston Hospital National Health Service (NHS) Trust, London, United Kingdom
[2]Leeds Institute of Health Sciences, University of Leeds, Leeds, United Kingdom

Corresponding Author:
Thomas Lorchan Lewis, BSc (Hons), MB ChB
Kingston Hospital National Health Service (NHS) Trust
Galsworthy Road
Kingston-Upon-Thames
London, KT2 7QB
United Kingdom
Phone: 44 7876453511
Email: tlewis@doctors.org.uk

Abstract

Background: One factor when assessing the quality of mobile apps is quantifying the impact of a given app on a population. There is currently no metric which can be used to compare the population impact of a mobile app across different health care disciplines.

Objective: The objective of this study is to create a novel metric to characterize the impact of a mobile app on a population.

Methods: We developed the simple novel metric, app usage factor (AUF), defined as the logarithm of the product of the number of active users of a mobile app with the median number of daily uses of the app. The behavior of this metric was modeled using simulated modeling in Python, a general-purpose programming language. Three simulations were conducted to explore the temporal and numerical stability of our metric and a simulated app ecosystem model using a simulated dataset of 20,000 apps.

Results: Simulations confirmed the metric was stable between predicted usage limits and remained stable at extremes of these limits. Analysis of a simulated dataset of 20,000 apps calculated an average value for the app usage factor of 4.90 (SD 0.78). A temporal simulation showed that the metric remained stable over time and suitable limits for its use were identified.

Conclusions: A key component when assessing app risk and potential harm is understanding the potential population impact of each mobile app. Our metric has many potential uses for a wide range of stakeholders in the app ecosystem, including users, regulators, developers, and health care professionals. Furthermore, this metric forms part of the overall estimate of risk and potential for harm or benefit posed by a mobile medical app. We identify the merits and limitations of this metric, as well as potential avenues for future validation and research.

(J Med Internet Res 2015;17(8):e200) doi: 10.2196/jmir.4284

KEYWORDS
mHealth; medical app; mobile phone; metric; risk assessment; medical informatics apps; population impact; mobile health; patient safety; mobile app

Introduction

Overview

The growth in popularity of health and medical apps for health care professionals and patients is widely recognized given their numerous successful uses in a number of health care domains, including clinical health care delivery, education, and health promotion [1-4]. However, a number of concerns regarding the reliability and accuracy of apps have arisen, leading to calls for some form of quality assessment [5-7].

Evaluating the quality of mobile apps is a notoriously difficult problem which currently has no standard solution. Ideally, every medical app should be evaluated and tested by a range of experts to ensure its suitability and applicability to medicine. In practice, this is impossible to achieve given the exponential growth in the app market, low barriers to entry, limited resources, and rapid pace of development [6]. A number of models have been

XSL•FO
RenderX

proposed to help clinicians, app developers, regulators, and commissioning bodies to assess the quality of mobile apps, although it remains to be seen which model has the most utility in practical terms [5,8,9].

A previous paper proposing a framework for risk assessment for mobile medical apps identified many components that increase the potential for harm [5]. The components included inherent factors to the app, such as functionality, content, complexity, and lack of a fail-safe. External factors included the app user, inappropriate usage, inadequate training, and the likelihood of an error being detected. Furthermore, this paper suggested that an important component of the potential harm caused by a medical app is the overall impact that a mobile app has on a given population [5]. Lewis et al noted, "Risk is proportional to the number of patients affected, so disease prevalence or similar indices of the number of people likely to be affected by an error need to be considered."

It follows that a less harmful app used by a large population could pose a greater overall population safety risk than a more harmful app used by a small population. Therefore, it is clear that there is a need to develop a metric that will assess the population impact of mobile medical apps and will allow subsequent comparison across different disciplines.

The Problem: Assessing the Impact of a Specific App

There are a limited number of options currently available to assess the impact of a mobile medical app on any given population; these are shown in Table 1 and are ranked in order of accuracy.

It is important for the various stakeholders to be able to estimate and compare the likely population impacts of specific apps for the reasons shown in Table 2.

It is critical to be able to assess the number of people at risk from an unsafe app at any given time and currently there is no clear method of assessing this. Our objective was to propose, develop, and model a simple metric that can be used to estimate and compare the likely impact of a specific app on a population.

Table 1. Models currently in use for assessing the impact of a mobile app on a given population.

Assessment tool[a]	Examples	Advantages	Disadvantages
Detailed app analytics	High level metrics such as active users, time spent on app, and ethnographic and epidemiological data	Gold standard in terms of data detail Would enable precise population impact to be measured	Large volumes of data Not currently practical Relies on app developers releasing crucial business information Developer bias
mHealth studies [10]	Numerous mHealth studies testing the validity of mobile apps for health care	Rigorous independent trials Often have detailed metrics available Often note the quality of an app	Often focus on one specific app Not many studies available Not easy to compare apps from different disciplines
Number of app downloads	Basic metric available from a number of sources	Can easily compare apps from different disciplines	Information not easily accessible Many users only download an app for trial purposes No information about how often an app is used No information about intended audience Often not reported accurately
Educated guesswork	N/A[b]	Minimal knowledge required to provide estimate	Not accurate or precise Wide observer bias

[a]Assessment tools are ranked in order of accuracy.

[b]Not applicable (N/A).

Table 2. Key reasons for use of population impacts of mobile apps by stakeholders.

Stakeholder	Reason for estimating app impact on population
Regulator	To estimate and compare the overall risks posed if the app is unsafe, and to decide on the appropriate regulatory measures
	To understand the potential for population benefit from effective apps
Guideline developer (eg, NICE[a])	To help understand the impact of an app from a public health perspective
App developer	To justify investment decisions
	To guide update strategy
App users	May use the population impact as a surrogate indicator for quality
Clinicians advising users about the app	May use the population impact as a surrogate indicator for quality
Health insurers and funding schemes	To understand the likely payback from approving reimbursement of the cost of the app
Health economists	As part of an estimate of cost effectiveness of the app
	Could utilize AUF[b] as part of their ranking algorithm
App stores	Surrogate marker for quality

[a]National Institute for Health and Care Excellence (NICE).

[b]App usage factor (AUF).

Methods

Identification of a Simple Metric: App Usage Factor

A broad literature search for existing metric systems in use was carried out. The two authors of this paper (TLL, JCW) searched for relevant papers with regard to their suitability for use when applied to medical apps, however, none were found. The search was expanded to look for metrics in use in other aspects of technology. We felt that there were some useful analogies outside health care; these included *passenger miles* for comparing CO_2 emissions from alternative transport methods, the *readership* of newspapers and magazines (number of sales × number of readers per paper), and common *website metrics*, which capture the number of unique visitors × time spent per visit.

We also brainstormed the criteria for a useful metric. In our view, a good metric should display the following characteristics:

1. Simple to calculate from readily available information

2. Reflects both the number of users and the frequency of use

3. Generates a single, understandable figure within the range of 0 to 10, despite the hugely varying number of users per app (at least a million-to-one ratio)

4. Relatively stable over time for each app

5. Can be used as a denominator for adverse incident reports

6. Makes intuitive sense to users in the same way as does the Richter scale, for example.

7. Has good interobserver agreement and reliability.

Our proposed metric, the app usage factor (AUF), is defined as the logarithm of the product of the total number of active users of a mobile app (A_U) with the median number of daily uses of the aforementioned app (D_U). The formula for calculating the metric can be seen in equation 1 as follows:

$$AUF = \log_{10}(A_U \times D_U) \quad (1)$$

There are a number of points worth considering that enable this metric to fulfil all the desired characteristics. Certain considerations for the measurement of A_U and D_U include geographic boundaries (eg, AUF could be global or country specific), operating system version (eg, AUF could differ for iPhone and Android platforms), and app version number, which would need to be specified or standardized in order to make a meaningful assessment of the AUF.

There are also specific temporal considerations when calculating A_U, D_u, and AUF, in particular, the following:

1. The figure for each variable should be assessed as "stable," for instance, not changed significantly over the past 30 days. This is because the use of apps by users can change extremely rapidly, for example, due to media hype surrounding the release of a new app.

2. The AUF for a specific app can change over time as the app ecosystem evolves. Specific issues that may be encountered include app or operating system updates, which may significantly affect the functionality of the app.

On the basis of the considerations above, it is reasonable to calculate the AUF for a specific app on a quarterly basis, at minimum, or 30 days following a major app update. Given the well-recognized speed of app updates, this may well require further refinement [11].

Validation of the App Usage Factor Metric Against Proposed Criteria

Overview

Any novel metric requires appropriate validation to ensure its suitability and accuracy for the proposed task. A series of computer simulation models were constructed that allowed specific app usage scenarios to be explored. The objective was to validate the AUF metric against the desirable metric criteria, thus confirming its suitability and applicability for practical use.

XSL•FO
RenderX

Three specific scenarios were modelled using Python [12], a high-level, general-purpose programming language, as follows:

1. Exploring stability of AUF as a function of A_u and D_u, including determination of metric limits

2. Simulated app ecosystem model

3. Temporal stability of AUF.

Exploring Stability of the App Usage Factor as a Function of Au and Du, Including Determination of Metric Limits

A 200×200 linear spaced grid was constructed in Python to simulate the behavior of the metric with calculated values of the AUF based on $0 < A_U < 100,000$ and $0 < D_U < 50$. A secondary iterative process was used to explore the limits of the AUF while still returning usable results. Specific positive and negative limits were identified.

Simulated App Ecosystem Model

Figures from a recent study suggest that there are now 20,000 health and medical apps available to download from the major online app stores [7]. In order to validate the new metric, Python was used to construct a simulated dataset with 20,000 "apps," each with a random number of daily "uses" and a random number of active "users." The values for A_U and D_U were randomly distributed using a negative exponential probability distribution; this is a continuous probability distribution which describes the time between events in a Poisson process (ie, a process in which events occur continuously and independently at a constant average rate [13]). This probability function was chosen to reflect the decreasing probability of an app being *both* widely used (A_U) *and* used multiple times per day (D_U). The probability distribution function for an exponential distribution is shown in equation 2.

$$P(x) = \lambda e^{-\lambda x} \quad (2)$$

An iterative process was used to identify values for lambda in order to identify suitable limits for the maximum number of daily activities, and the maximum number of active users [14]. This was determined to be 0.00001 and 0.4 for A_U and D_U, giving maximum numbers of 1 million users and 30 uses per day, respectively.

Temporal Stability of the App Usage Factor

In order to assess the temporal stability of the AUF, the behavior of a single app was modelled as a function of time. The strength of the logarithmic component of the AUF is its ability to act as a damping system to external ecosystem factors (eg, media hype). The aim was to show that the AUF would not change dramatically in response to these factors. In order to test this, a series of external ecosystem factors would be applied to the simulation to observe how the AUF changed. Key criteria for the simulation were as follows:

1. D_U calculated to be a random float value between minimum and maximum values of D_U ($D_{U\ MIN}$ and $D_{U\ MAX}$, respectively) according to a uniform probability distribution each day

2. A_U calculated to increase/decrease by x users each day, where x is a float value determined by a uniform probability distribution between minimum and maximum values of A_U ($A_{U\ MIN}$ and $A_{U\ MAX}$, respectively)

3. AUF calculated according to values of A_U and D_U as described by the metric and modelled as a function of time for a period of 2 years.

Specific functions were applied to the model, which intended to simulate the following external app ecosystem events: initial market launch, positive media publicity, negative media publicity, and app/operating system updates. The AUF for the simulated app was then plotted as a function of time to analyze temporal behavior. The values for initial A_U ($A_{U\ INITIAL}$), $A_{U\ MIN}$, $A_{U\ MAX}$, $D_{U\ MIN,}$ and $D_{U\ MAX}$ for each external ecosystem event are shown below in Table 3.

Table 3. Initial data used to model the characteristics of the AUF as a function of time for a single mobile app.

External ecosystem event	Day number	$A_{U\ MIN}$ [a]	$A_{U\ MAX}$ [b]	$D_{U\ MIN}$[c]-$D_{U\ MAX}$[d]
Initial market launch	1[e](No. of users initially set at 50)	-50	50	10-20
Daily market fluctuation	All days other than those below	-50	50	10-20
Positive media publicity	100-110	50	500	10-20
Negative media publicity	350-360	-500	50	10-20
App version/operating system update	501[f](No. of users reset to 500)	-50	50	10-20
Users upgrade to latest version	500-650	-20	250	10-20

[a]Range for minimum number of active users of a mobile app ($A_{U\ MIN}$).

[b]Range for maximum number of active users of a mobile app ($A_{U\ MAX}$).

[c]Range for minimum median number of daily uses of an app ($D_{U\ MIN}$).

[d]Range for maximum median number of daily uses of an app ($D_{U\ MAX}$).

[e]Initial number of active users of a mobile app ($A_{U\ INITIAL}$) on day 1 (initial market launch)=50.

[f]At day 501, the number of active users was reset to 500 to simulate app version/operating system update.

Results

Stability of the App Usage Factor as a Function of AUand DU

The results of our model shown in Figure 1 highlight the value for AUF (contour lines) as a function of A_U and D_U. The metric remains stable (ie, AUF > 0) provided $A_U \times D_U$ is greater than 1. It is reasonable to assume that if $A_U \times D_U$ were less than 1, then the specific app would be rarely used, if at all. In practical terms, this relates to a scenario where an app is used regularly once a month by 50 people or less, which was deemed to be an acceptable minimum standard.

Figure 1. A contour plot illustrating the stability of the app usage factor as a function of Au and Du, including determination of metric limits.
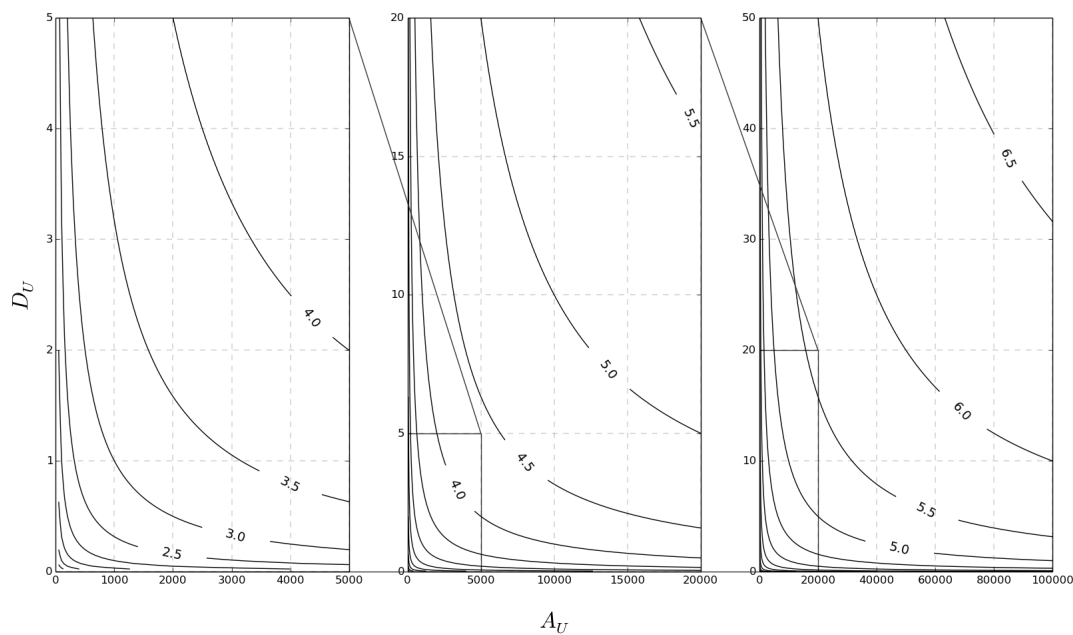

Simulated App Ecosystem Model

The input data can be seen on the left in Figure 2, with the relative frequency of both A_U and D_U plotted as histograms on the right. Each data point represents an individual mobile app with an independent, randomly assigned A_U and D_U. The AUF was then calculated and plotted as a histogram against frequency as shown in Figure 3. As a result of the logarithmic scaling factor, each unit increase in AUF represents a factor of 10 for impact on the population, similar to the Richter scale.

Apps with a similar AUF can be considered to have a comparable population impact to each other, while simultaneously giving a useful indication of the scale of users affected (see Table 4). Distribution of the sample results can be seen in Figure 3. The interquartile range for AUF was calculated to be 4.45 to 5.45. The mean AUF was 4.90, with the standard deviation calculated to be 0.78.

Figure 2. A combined scatterplot (input data, left) and histogram (relative frequency of both Au and Du, right) showing the initial sample dataset of 20,000 mobile medical apps.
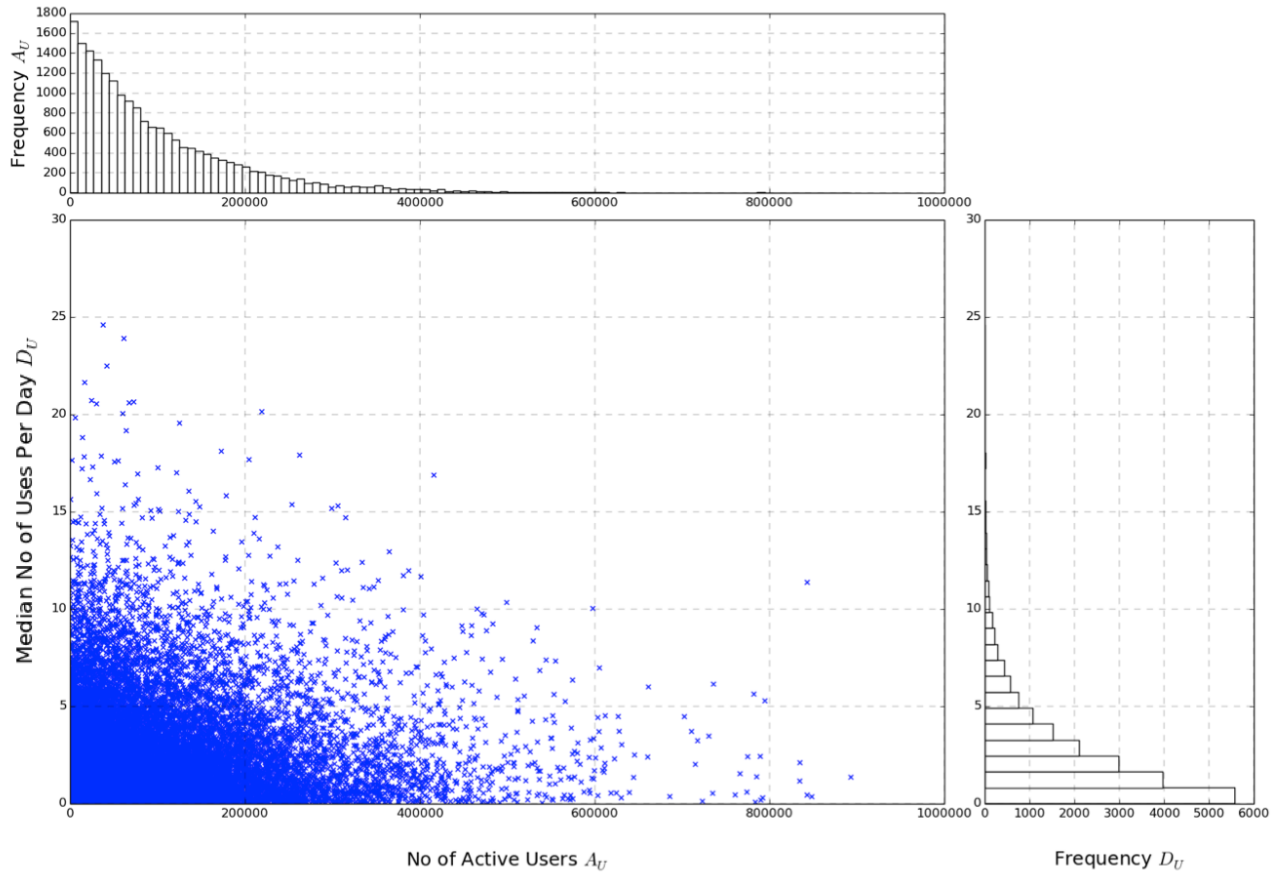
Figure 3. A histogram showing the frequency distribution of the app usage factor for the sample dataset of 20,000 simulated mobile medical apps, including mean and standard deviation for the data.
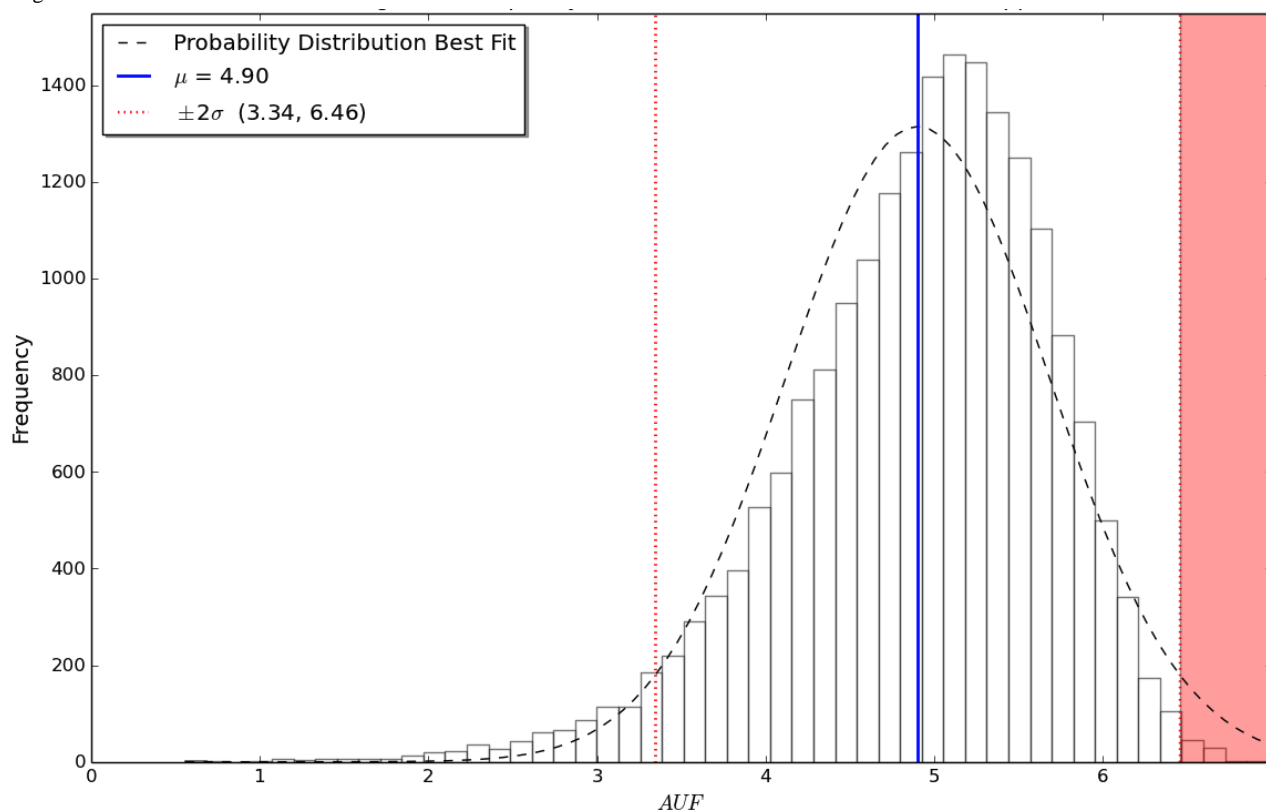
Temporal Stability of the App Usage Factor

The temporal simulation shown in Figure 4 shows that the AUF will tend toward a relatively stable state despite market perturbations. Minor daily fluctuations in A_U and D_U are effectively dampened by the addition of the logarithmic factor. Temporal simulations carried out without the logarithmic factor show an increased sensitivity to small changes in A_U and D_U, which lead to an overall decrease in stability over time. Our simulations suggest that there is a lag time present between a major app ecosystem perturbation and the corresponding change in AUF. This is likely to represent the time taken for information to reach the affected user base and is therefore dependent on the number of users and the magnitude of the market perturbation. It follows that a small market perturbation affecting a small number of users is unlikely to significantly affect the AUF. On the basis of this simulation, it is appropriate to delay measuring AUF for 30 days after any minor market perturbation and 80 to 100 days after a large market perturbation (eg, app launch, operating system update).

Figure 4. A graph showing app usage factor as a function of time for a single mobile app which is subject to a number of simulated app ecosystem events.
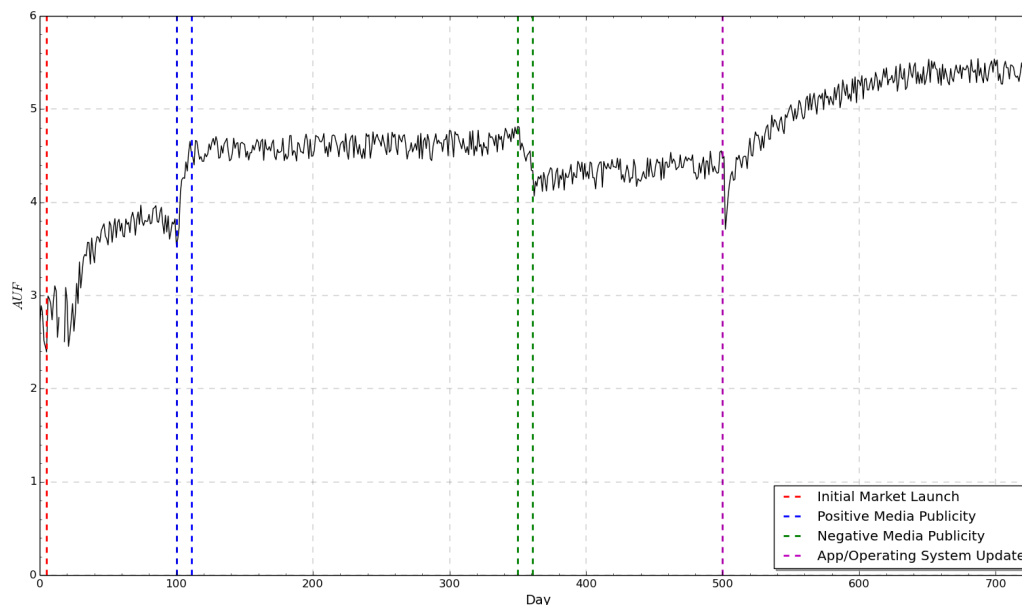
Discussion

Merits and Limitations of the App Usage Factor Metric

Use of the AUF as a metric to assess the population impact of a mobile medical app has many potential benefits for health care professionals, developers, and regulators alike. People who use this metric should be able to make a comparison of the AUF with real-world usage of a particular app, as can be seen in Table 4.

The biggest limitation of this metric is obtaining the two key pieces of information: How many active users there are (A_U) and how many times a day the app is used (D_U). It is not possible to accurately calculate the AUF without approximate figures for these variables. Identification of A_U and D_U could be facilitated through the following means:

1. Voluntary release by app developers/app store companies to an independent body where possible, perhaps as part of a self-certification process. The AUF could be calculated by developers and release of this data does not in itself release confidential information.

2. Compulsory release as part of a formal regulatory process, for example, to gain Food and Drug Administration (FDA) regulation approval.

3. Survey of a target population with subsequent data extraction and extrapolation of log files associated with app usage.

There are a number of further considerations of the app usage factor for risk assessment of mobile apps, particularly when utilizing AUF to estimate population impact of an app.

Risk-based regulatory models such as those utilized by the FDA [15] and the Medicines and Healthcare products Regulatory Agency (MHRA) [16] could potentially target apps with a high AUF (ie, a large population impact) purely on the basis of potential negative impact to a population, as suggested in the red highlighted area in Figure 3. Previous risk assessment analysis of mobile apps identified that approximately 0.5% of apps require this formal regulatory assessment. In our opinion, it is therefore reasonable to identify the apps which have the highest population impact and assess their safety on a case-by-case basis. Identification of these apps is accomplished by identifying all apps that fall more than 2 standard deviations from the mean AUF. In our sample data, this equated to 82 out of the 20,000 simulated apps, a feasible proportion of 0.41% of the apps which could require assessment pending formal regulation; these are highlighted in red in Figure 3. In our simulation, this equates to an app with an AUF greater than 6.46.

A limitation of the AUF in estimating the potential risk of a mobile app is when the number of active users is high within a potentially small user base. This is illustrated by the following thought experiment. A potentially harmful app used regularly once a day by 600 users out of a population of 800 users has a higher chance of causing harm to this user group than the AUF of 2.8 for this scenario alone would suggest. This thought experiment confirms that the AUF is a measure of *population* impact and there are several contextual and other factors that contribute to overall risk posed by a medical app [2].

Table 4. Equivalent population impact of an app based on its corresponding AUF.

App usage factor (AUF)	Equivalent active user daily actions ($A_U{}^a \times D_U{}^b$)
6	1,000,000
5	100,000
4	10,000
3	1000
2	100

[a]Number of active users of a mobile app (A_U).

[b]Median number of daily uses of an app (D_U).

Conclusion

A key component when assessing app risk and potential harm is understanding the potential population impact of each mobile app. Our new metric would have many potential uses for a wide range of stakeholders in the app ecosystem, including users, regulators, developers, and health care professionals. Furthermore, this metric forms part of the overall estimate of risk and potential for harm or benefit posed by a mobile medical app [2]. We developed and explored the characteristics of a novel but simple, easily calculated metric to assess the likely population impact of a medical app using a sample database of 20,000 apps modelled using a computer simulation. This modelling showed that our proposed metric, AUF, remained stable over time and at extremes of user numbers and daily usage rates, thereby confirming its suitability for further testing in a health care context. We are confident that using this metric will help the population impact of a specific app to be estimated and compared with similar apps. It is important to note that AUF forms but one component of the overall risk and harm potential posed by a specific app. Users should take the AUF into consideration alongside inherent and external risk factors when deciding whether to use an app in clinical practice. For now, the next stage in the validation process is to calculate the app usage factor for a number of health and medical apps using actual usage and population data.

Acknowledgments

Conflicts of Interest

TLL is a writer and editor for iMedicalApps.com, a website dedicated toward providing news on the integration of mobile technology into medical care and the reviewing of medical apps for mobile devices. He does not consult nor receive reimbursement from app developers or creators. JW has no competing interests to declare.

References

1. Lupton D. Apps as artefacts: towards a critical perspective on mobile health and medical apps. Societies 2014 Oct 29;4(4):606-622. [doi: 10.3390/soc4040606]
2. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
3. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
4. Lewis TL, Vohra RS. Smartphones make smarter surgeons. Br J Surg 2014 Mar;101(4):296-297. [doi: 10.1002/bjs.9328] [Medline: 24496760]
5. Lewis TL, Wyatt JC. mHealth and mobile medical apps: a framework to assess risk and promote safer use. J Med Internet Res 2014;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
6. Misra S, Lewis TL, Aungst TD. Medical application use and the need for further research and assessment for clinical practice: creation and integration of standards for best practice to alleviate poor application design. JAMA Dermatol 2013 Jun;149(6):661-662. [doi: 10.1001/jamadermatol.2013.606] [Medline: 23783150]
7. Aungst TD, Clauson KA, Misra S, Lewis TL, Husain I. How to identify, assess and utilise mobile medical applications in clinical practice. Int J Clin Pract 2014 Feb;68(2):155-162. [doi: 10.1111/ijcp.12375] [Medline: 24460614]
8. Lewis TL. A systematic self-certification model for mobile medical apps. J Med Internet Res 2013;15(4):e89 [FREE Full text] [doi: 10.2196/jmir.2446] [Medline: 23615332]

9.   Albrecht U. Transparency of health-apps for trust and decision making. J Med Internet Res 2013;15(12):e277 [FREE Full text] [doi: 10.2196/jmir.2981] [Medline: 24449711]

10.   Frisbee K. The King's Fund. 2014 Oct 02. Assessing the impact of mobile health apps URL: http://www.kingsfund.org.uk/audio-video/kathleen-frisbee-assessing-impact-mobile-health-apps [accessed 2015-08-10] [WebCite Cache ID 6aiA0XyQ7]

11.   van Velsen L, Beaujean DJ, van Gemert-Pijnen JE. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak 2013;13:23 [FREE Full text] [doi: 10.1186/1472-6947-13-23] [Medline: 23399513]

12.   Python.: Python Software Foundation URL: https://www.python.org/ [accessed 2015-08-10] [WebCite Cache ID 6ai9o07V0]

13.   Weisstein EW. Wolfram MathWorld. Exponential distribution URL: http://mathworld.wolfram.com/ExponentialDistribution.html [accessed 2015-08-11] [WebCite Cache ID 6ajaU1ItA]

14.   Payne KB, Wharrad H, Watts K. Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]

15.   US Food and Drug Administration. Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff. Silver Spring, MD: US Department of Health and Human Services, Food and Drug Administration; 2015 Feb 09. URL: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf [accessed 2014-12-10] [WebCite Cache ID 6ai9gIwNc]

16.   Medicines and Healthcare Products Regulatory Agency (MHRA). 2010 May 12. Medical Device Technology Forum on the use of software as a medical device URL: http://webarchive.nationalarchives.gov.uk/20141205150130/http://www.mhra.gov.uk/Howweregulate/NewTechnologiesForums/DevicesNewTechnologyForum/Forums/CON084987 [accessed 2015-08-11] [WebCite Cache ID 6ajaM8kA2]

Abbreviations

AU: number of active users of a mobile app
AUINITIAL: initial AU
AU MAX: maximum value of AU
AU MIN: minimum value of AU
AUF: app usage factor
DU: median number of daily uses of an app
DU MAX: maximum value of DU
DU MIN: minimum value of DU
FDA: Food and Drug Administration
MHRA: Medicines and Healthcare products Regulatory Agency
MRC: Medical Research Council
N/A: not applicable
NICE: National Institute for Health and Care Excellence