Research Note

Automated Characterization of Mobile Health Apps' Features by Extracting Information From the Web: An Exploratory Study

Alessia Paglialonga,[a] Massimo Schiavo,[b] and Enrico Gianluca Caiani[a,b]

Purpose: The aim of this study was to test the viability of a novel method for automated characterization of mobile health apps.
Method: In this exploratory study, we developed the basic modules of an automated method, based on text analytics, able to characterize the apps' medical specialties by extracting information from the web. We analyzed apps in the Medical and Health & Fitness categories on the U.S. iTunes store.
Results: We automatically crawled 42,007 Medical and 79,557 Health & Fitness apps' webpages. After removing duplicates and non-English apps, the database included 80,490 apps. We tested the accuracy of the automated method on a subset of 400 apps. We observed 91%

accuracy for the identification of apps related to health or medicine, 95% accuracy for sensory systems apps, and an average of 82% accuracy for classification into medical specialties.
Conclusions: These preliminary results suggested the viability of automated characterization of apps based on text analytics and highlighted directions for improvement in terms of classification rules and vocabularies, analysis of semantic types, and extraction of key features (promoters, services, and users). The availability of automated tools for app characterization is important as it may support health care professionals in informed, aware selection of health apps to recommend to their patients.

The rapid growth of mobile health (mHealth) technology is creating an entirely new research area and opportunities for better care. Also in the area of hearing health care (HHC), mHealth is becoming increasingly popular (Bright & Pallawela, 2016; Paglialonga, Pinciroli, & Tognola, in press; Paglialonga, Tognola, & Pinciroli, 2015; Yousuf Hussein, Swanepoel, Mahomed, & Biagio de Jager, 2018). In this context, the wide adoption of smartphones in the general population and the proliferation in number of health-related apps bring unprecedented opportunities. Several actors are involved (including patients, citizens, and health care professionals), across a variety of medical specialties and clinical applications. Potential benefits include,

for example, promotion of preventive behaviors and health monitoring, enhanced patient–doctor interaction, improved service delivery in resource-limited settings, patient empowerment, and patient-centered care (Kim & Lee, 2017; Paglialonga, Mastropietro, Scalco, & Rizzo, in press; Scherer, Ben-Zeev, Li, & Kne, 2017). However, in this rapidly evolving field, some unforeseen challenges are also emerging. For example, concerns about data protection, risks related to app safety and misuse, and poor regulation, as well as a substantial lack of systematic methods for app identification, characterization, and quality assessment (BinDhim & Trevena, 2015; Misra, Lewis, & Aungst, 2013; Paglialonga, Lugo, & Santoro, 2018), could hinder the use of these novel tools.

Enabling Informed, Aware Adoption of Health Apps

A recent report estimated that over 325,000 health apps were available on major stores (Apple, Google Play, and Windows store combined) in 2017 (Research2Guidance, 2017). In this very large market, finding the right app for a specific need can be challenging. Similarly, it may be difficult to identify the relevant features of an app before downloading it. These obstacles might limit the users' confidence in these tools and, in turn, the potential benefits associated with their utilization. To this purpose, a new area

[a]Consiglio Nazionale delle Ricerche, Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni, Milan, Italy
[b]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy

Correspondence to Alessia Paglialonga: alessia.paglialonga@polimi.it; alessia.paglialonga@ieiit.cnr.it

of research is emerging, focusing on the development of tools and methods to provide accurate, meaningful information to potential app users and thus to enable informed, aware adoption of health apps (Aungst, Clauson, Misra, Lewis, & Husain, 2014; Grundy, Wang, & Bero, 2016; Jungnickel, von Jan, & Albrecht, 2015; Paglialonga, Lugo, et al., 2018; Paglialonga, Pinciroli, Tognola, Barbieri, et al., 2017; Wyatt et al., 2015). It is worth noting that the term *potential users* for health apps includes not only the individuals directly interested in using the app (e.g., the patients or their caregivers) but also those who can direct toward the use of an app, such as health care professionals. In the ongoing digital health transformation, physicians and health care professionals can play a pivotal role to identify the most suited mobile apps to be recommended to patients (Aungst et al., 2014).

Professional advice can be the key to enable effective use of mobile solutions by patients and their families. However, no clear strategy has emerged on how providers should evaluate and recommend health apps to patients (Singh et al., 2016). In addition, important barriers exist, and only a small proportion of physicians currently tend to recommend mHealth apps to their patients, also including early adopter countries (Zhang & Koch, 2015). This is related to various factors, for example, how to integrate apps into the clinical workflow, how to motivate patients to download and use them, and, especially, how to choose apps to be prescribed and how to assess their quality (Terry, 2015; Wyatt et al., 2015). For clinicians and, in general, for potential app users, it is difficult to find comprehensive and reliable resources able to guide them through informed app adoption and use. In fact, searching capabilities on app stores allow the user to focus on a specific category (i.e., Health & Fitness [H&F]) in which the order of recommended apps is determined by the number of downloads, with no relation with the accuracy of the app or its potential clinical relevance. As an alternative, searching for a specific term (i.e., hearing), it provides a long list of thumbnails that include congress-related apps, educational materials, games, apps that are not related to the term, and also some that could be useful to be prescribed, but to obtain more detailed information, there is a need to go through the specific app page by manually visualizing it, making this process impractical. Related to this, studies among different target groups (medical doctors, medical students, and citizens) showed that potential app users are concerned about this overload of health apps available on the market and how to verify their quality and filter out what is not relevant (Aungst et al., 2014; Franko & Tirrell, 2012; Paglialonga, Lugo, et al., 2018; Payne, Wharrad, & Watts, 2012). These concerns can translate into barriers to app adoption as, before health care providers or organizations can recommend an app to patients, they need to be confident about the app's function, effectiveness, and quality (Boudreaux et al., 2014) to avoid possible liability.

In this evolving context, the development of methods able to provide meaningful information about health apps would be thus particularly valued to improve knowledge and awareness among physicians and health care professionals who need to choose apps to recommend to patients. Classification needs to be effective, to allow accurate app identification in a given topic area. Characterization needs to be informative and able to highlight the core features of an app, to whom it is addressed, and a better specification of its clinical interest (Paglialonga, Lugo, et al., 2018). With this goal in mind, we recently developed a new method to characterize and assess apps for HHC: the At-a-Glance Labeling for Features of Apps for Hearing Health Care (ALFA4Hearing). This is a descriptive model to characterize apps against a core set of 29 features, which can be coded easily and with negligible interrater variability (Paglialonga, Pinciroli, & Tognola, 2017; Paglialonga, Pinciroli, et al., in press). The features included in the model cover five general domains: the app's promoters, the offered services, the relevant implementation of features, the targeted users, and general descriptive information. As such, the model can be used to describe any given app for HHC, regardless of the operating system and mobile platform. It has proved to be useful to review subsamples of apps, to create descriptive pictures of the app market, and to assess the emerging trends, challenges, and potential opportunities in the field (Paglialonga, Pinciroli, et al., in press). In general, descriptive models such as the ALFA4Hearing mentioned here are useful to characterize apps and extract features that are related to app functionalities, offered services, target groups, and quality. Several approaches have been proposed for app characterization and assessment of quality components (reviewed by Paglialonga, Lugo, et al., 2018), but their use needs extensive work as single apps need to be downloaded and reviewed manually. Considering the very high number of apps on the market and their frequent updates, it would be important to explore the feasibility of using automated methods to classify and characterize apps potentially in real time.

Need for Automated Methods for App Characterization

The market of health apps grows rapidly at an estimated rate of about 100,000 per year (Research2Guidance, 2017). In addition, apps are updated regularly, typically at least once after every major update of the operating systems. Therefore, recommendations about apps can hardly be complete and up-to-date if manually generated. Given the time needed to review apps and their features, characterizing apps manually requires a very high amount of time and effort. For these reasons, it is important to explore possible automated methods to fill out descriptive models in real time as this would help to keep pace with the very rapid development and update of apps on the market (Paglialonga, Riboldi, Tognola, & Caiani, 2017). The long-term aim of our research is to develop automated methods to extract meaningful information about apps' features and quality components directly from the web (e.g., vendor markets, developers' websites, app repositories). As a first step, our short-term aim was to develop

the basic modules of a novel automated method, based on text analytics, to extract relevant information directly from the unstructured text description reported on the apps' webpages in the app vendor markets. To assess the feasibility of this automated approach, we focused on the extraction of one general feature, that is, the app's medical specialty (e.g., HHC, neurology, nutrition), and we analyzed all the apps available in the Medical (M) and H&F categories on the U.S. iTunes app store.

Method

App Database Creation

Figure 1 outlines the proposed automated method for app characterization. All the functions in the modules were developed in Python by using PyCharm (JetBrains, Prague, Czech Republic), a cross-platform integrated development environment. The first module (see Figure 1a) generates the app database.

- URL builder: The function "URL builder" crawls the app store to extract the URLs (uniform resource locators, i.e., the web addresses) of the webpages of all the apps in the M and H&F categories. First, by taking advantage of the URLs' predetermined structure, the function generates dynamically the URLs of the app store webpages where the names and hyperlinks of all the available apps are listed alphabetically. For example, the URL http://itunes.apple.com/us/genre/ios-medical/id6020?mt=8&letter=T&page=3#page locates the third page ("&page=3#page") of the list of apps whose name begins with the letter "T" ("&letter=T") in the M category ("genre/ios-medical/") in the U.S. store ("/us/"). In this way, the function is able to access every page in the list, for each initial, and for each category. From each of these pages, the hyperlink of each app in the list is extracted, and by using the library "REQUESTS," the app's webpage on the store is accessed and the HTML (HyperText Markup Language) source code is extracted.

- HTML analyzer: The HTML is the set of markup symbols or codes used for display on a World Wide Web browser page (HTML Working Group team at

Figure 1. Outline of the proposed automated method for app characterization. (a) App database creation. The function "URL builder" crawls the app store in the M and H&F categories by dynamically building the URLs of the app store webpages where the apps' names and hyperlinks are listed alphabetically and accesses each app's webpage. In each app's webpage, the function "HTML analyzer" parses the HTML source code and extract apps' attributes to build the app database. (b) Feature extraction. The function "database cleaner" preprocesses text data in the database by removing, based on the app description, apps that are not in English and by converting HTML into plain text by using ASCII (7 bits) characters. Then, the function "feature extractor" uses text analytics to extract UMLS terms and, based on the MeSH hierarchical structure, to characterize apps' features (in this pilot study, the medical specialty). M = Medical; H&F = Health and Fitness; UMLS = Unified Medical Language System; MeSH = Metathesaurus and Medical Subject Headings.

the World Wide Web Consortium, 2017). An HTML code is well structured, and its structure and tags do not depend on the app; therefore, it can be parsed by using regular expressions (REs). REs are used in programming to look for a specific pattern in text (Aho & Ullman, 1992). The function "HTML analyzer" (see Figure 1a) parses the HTML code, which has the same format for each app on the store, by using the python library "RE" and extracts the apps' attributes to create the database. For example, the RE "</h1 > \s* < h2 > By(.*?)</h2>" can be used to look for the developer's name in the HTML code, and the RE "<h1 itemprop = "name" > (.*?)</h1>" can be used to extract the app's name. For a comprehensive description of apps to be used in this pilot as well as in future studies, 16 attributes were extracted from each app's webpage and stored in a Comma Separated Values file. Specifically, the database included the following attributes: app ID (a unique identifier on the store), name, description (i.e., the unstructured text analyzed in this first study), version, developer's name, developer contacts, last update date, device compatibility, iOS compatibility, number of ratings, average ratings, reviews' content, price (in USD), size (in megabytes), URL, and timestamp (i.e., the date and time of webpage access, set by the function "HTML analyzer"). As the app ID represents a unique identifier of the app on the store, it was used here as the primary key to avoid duplicates in our app database. Duplicates are inherently present on the app store because some apps are listed in both M and H&F categories, some have more than one name, and some can be found, with the same name, under more than one initial letter.

Automated Feature Extraction

The second module (see Figure 1b) extracts the apps' features.

• Database cleaner: The function "database cleaner" processes text data in the database (specifically, the app description) to remove apps with empty, very short (below 20 characters), or non-English description. Then, the function converts HTML into plain text (ASCII 7 bits). Identification of apps in languages other than English was necessary as it was observed that several apps on the U.S. store were claimed to be in English but their description was provided in a different language. Removal of these apps was needed as our method uses tools for English text analytics. Language detection was performed by using the Google Cloud Translation API Client Library for Python tool (Google LLC, Mountain View, CA), specifically a port of Google's language detection library to Python.

• Feature extractor: The second function in this module, feature extractor, is the core of the proposed automated approach and uses text analytics to characterize the apps' features. As, in this pilot study, the focus was on the extraction of the app's medical specialty, a text analytics tool able to extract information related to medical concepts was chosen: MetaMap (National Library of Medicine [NLM], Bethesda, MD), a program developed for automatic indexing of biomedical literature at the NLM. MetaMap is a highly configurable program that uses a knowledge-intensive approach based on symbolic natural language processing and computational linguistic techniques (Aronson, 2001). MetaMap was used in this study as it maps unstructured text (as in the apps' descriptions) to the Unified Medical Language System (UMLS) Metathesaurus and Medical Subject Headings (MeSH). MeSH is the NLM's controlled vocabulary thesaurus, which consists of sets of terms naming descriptors in a hierarchical structure that allow searching at various levels of specificity. The function "feature extractor" uses MetaMap to extract, from the unstructured text in the apps' description, UMLS terms and the related scores that represent the probability of correct interpretation of the sentence (Aronson, 2001). For improved efficacy, 49 of the 129 available semantic types were selected as relevant (e.g., body location or region; body part, organ, or organ component; food), whereas 80 were excluded as not relevant or misleading (e.g., amphibian, bird, carbohydrate sequence). Then, from each retrieved term, the function browses the MeSH hierarchical structure and maps it into one or more medical specialties: As an example, the term *auditory disease* will be mapped into HHC and neurology. On the basis of this mapping, the function computes for each specialty a relevance score (RS) as the number of terms retrieved in that specialty multiplied by their average associated score, normalized in the range of 0–1. Then, the identified medical specialties are ranked based on the RS, and each app is matched with one or more of them by selecting those ranked in the top 10% of the RS. If no UMLS terms are retrieved from the app's description, or if only one term with a score lower than 0.8 is retrieved, then the app is characterized as not related to health or medicine. These cutoff thresholds were set based on a preliminary analysis of a subset of 250 apps that allowed selecting the most reliable terms in mapping and the highly relevant medical specialties in classification.

Testing Set for Manual Characterization

To assess the performance of the proposed approach, a testing set of apps was manually reviewed, to be compared with the results of the automated classification. A subset of 400 apps from the M and H&F categories on the U.S. iTunes app store was randomly selected from the app database by generating 400 distinct random numbers from a uniform statistical distribution. The apps were manually reviewed by two researchers (A. P. and M. S.) who extracted the following features blinded to each other, using predetermined categories: (a) medical specialties,

(b) promoters, (c) offered services, and (d) target users. In case of different classifications, the results of this manual review were discussed until consensus was reached. More than one category was coded as relevant whenever needed. Coding of features was based on categories conceptualized in reference standards (European Union of Medical Specialists, n.d.) and in previous studies (Paglialonga, Pinciroli, & Tognola, 2017; Paglialonga, Riboldi, et al., 2017). Specifically:

- Medical specialty: across specialties (i.e., apps generally related to medicine, medical education, nursing, and health care rather than to one or more specialties), cardiology, dentistry, dermatology, diabetes care, emergency medicine, fitness & wellness, gastroenterology, gynecology and obstetrics, immunology and endocrinology, mental health and neurology, nutrition, oncology, pediatrics, physiatry and orthopedics, sleep and respiratory care, sensory systems health care (including HHC, speech and language rehabilitation [S&L], vestibular medicine, and vision health care), surgery, and urology

- Promoters: companies (e.g., device manufacturers, drug companies, software houses); fitness & wellness providers; health care providers (i.e., clinics, pharmacies, insurance companies); individual developers; patients, families, or citizens' associations; professional, scientific, and educational institutions; public health and governmental institutions; and publishers and communication agencies

- Offered services: assistive tool, clinical support tool, education and information, electronic health record/personal health record, practicing and tracking, purchase, service finder, telemedicine, and testing and self-testing

- Target users: citizens, patients, physicians and health care professionals, and families and significant others

Results

App Database

By using the function "URL builder," 42,007 M and 79,557 H&F apps' webpages were crawled on the U.S. iTunes app store as of May 31, 2017. Some apps (i.e., 37 M and 68 H&F) had empty or very short descriptions (below 20 characters), and a significant number were described in non-English languages (i.e., 11,397 M equal to 27% and 18,382 H&F equal to 23%). After removal of these apps and of duplicates across categories, a database of 80,490 unique apps was obtained: 19,383 (24.1%) in the M category, 49,917 (62.0%) in the H&F category, and 11,190 (13.9%) in both categories.

The testing set for manual review included 400 apps randomly selected from the app store, resulting in 100 (25.0%) from the M category, 245 (61.2%) from the H&F category, and 55 (13.8%) mapped in both categories.

Manual Characterization

When coding for medical specialty, a not negligible number of apps (i.e., 59/400, 14.8%) were found to be not related to health or medicine (e.g., entertainment, games, business apps), although they had been originally classified as M (10), H&F (44), or both (5) on the app store by their developers. The distribution of the remaining 341 apps across medical specialties is reported in Figure 2a.

- Medical specialties: Of these 341 apps, 139 (40.8%) were classified as Fitness & Wellness (including fitness, wellness, sports, lifestyle management, yoga, stress relief, massages, and alternative medicine), 75 (22.0%) were mapped across specialties, 20 (5.9%) were related to nutrition, 17 (5.0%) were related to sensory systems health care (HHC: three apps, S&L: six apps, vestibular medicine: one app, vision health care: seven apps), 14 (4.1%) were related to cardiology, 11 (3.2%) were related to gynecology and obstetrics, 10 (2.9%) were related to emergency medicine, eight (2.3%) were related to mental health and neurology, seven (2.1%) were related to dentistry, six (1.8%) were related to diabetes care, six were related to oncology, six were related to physiatry and orthopedics, five (1.5%) were related to dermatology, four (1.2%) were related to surgery, four were related to sleep and respiratory care, three (0.9%) were related to urology, two (0.6%) were related to gastroenterology, two were related to pediatrics, and two were related to immunology and endocrinology.

- Promoters: Coding for the app promoters showed that companies (e.g., device manufacturers, drug companies, software houses) were involved in the development of a large subset of apps, that is, 75 of 341 apps (22.0%). Health care providers and fitness & wellness providers also played an important role with 52 (15.2%) and 53 (15.5%) apps, respectively. Professional, scientific, and educational institutions supported the development of 25 apps (7.3%); publishers and communication agencies promoted 20 apps (5.9%); patients, families, or citizens' associations were involved in the development of 11 apps (3.2%); and public health and governmental institutions were involved in eight apps (2.3%), whereas 101 apps (29.6%) were promoted by individual developers, that is, developers not related to any of the above categories.

- Services: In terms of services, 132 apps (38.7%) offered practicing and tracking functionalities (including training and rehabilitation), 121 apps (35.4%) had tools for education and information, 31 apps (9.1%) provided purchasing services (e.g., drugs, devices), 23 apps (6.7%) represented clinical support tools (e.g., decision tools, clinical data analysis tools, patient management services for health care professionals), 15 apps (4.4%) were for testing (including self-testing), seven apps (2.1%) were assistive tools for patients, six apps (1.8%) included services related to the electronic health record or personal health record, four apps (1.2%) were

Figure 2. Distribution of apps across medical specialties. (a) Distribution of apps across medical specialties in the testing set as obtained by manual characterization. (b) Percentages of apps correctly categorized across medical specialties as obtained by automated characterization (*N* = 341 apps).

(a)

Distribution of apps across Medical Specialties

(b)

Automated Characterization of apps across Medical Specialties
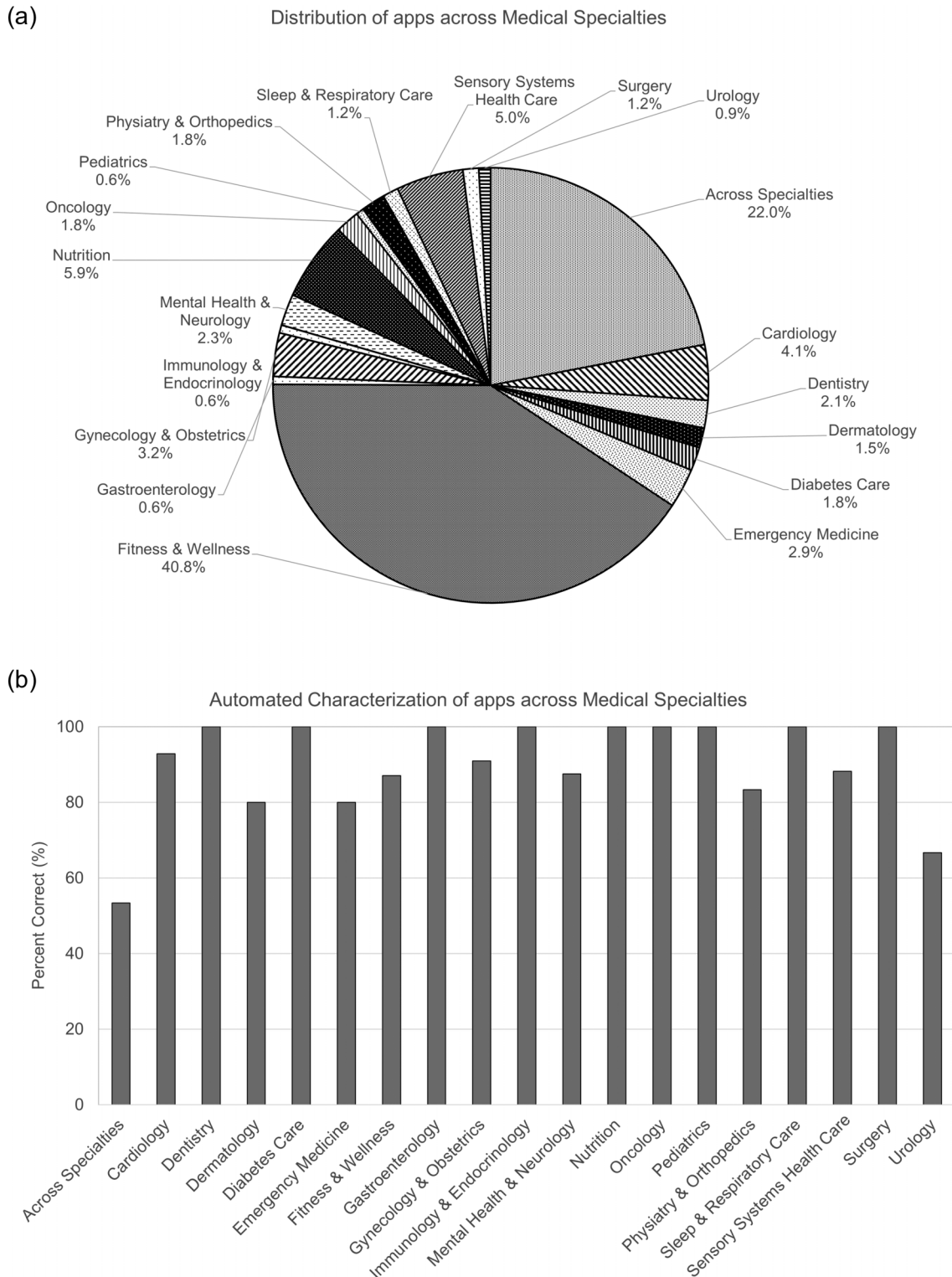
telemedicine tools, and two apps (0.6%) offered service finder functionalities.

- Target users: The analysis of the target users for apps in the testing set showed that 78 of 341 apps were related to more than one user group (typically, citizens and patients, patients and families/significant others, and patients and physicians/health care professionals). The primary target user group was citizens in 174 apps (51.0%), patients in 93 apps (27.3%), physicians/health care professionals in 71 apps (20.8%), and families/significant others in three apps (0.9%). Overall, by counting primary and secondary user groups, we found that 209 of 341 apps could be used by citizens; 111 apps, by patients; 80 apps, by physicians/health care professionals; and 19 apps, by families/significant others.

Apps for Audiology

A closer look at the apps related to audiology (10 apps in the testing set, i.e., HHC, S&L, and vestibular medicine combined) showed that four were promoted by individual developers; three, by companies (specifically, by hearing aid manufacturers); one, by public health and governmental institutions (a communication app); one, by publishers and communication agencies (a reference guide for vertigo); and one, by patients, families, or citizens' associations (a communication and speech learning app for children with special needs). As regards the offered services and targeted groups, three of five were assistive tools (specifically, communication tools) for patients, three were for education and information about hearing aids for patients with hearing loss, two were for testing (specifically, a hearing screening test for adults and the Infant–Toddler Meaningful Auditory Integration parent report), and two were for hearing rehabilitation and S&L practicing.

Automated Characterization

Automated analysis of apps by using the function "feature extractor" correctly characterized 304 of 341 apps as related to health or medicine and 33 of 59 as not related to health or medicine (sensitivity = 89%, specificity = 56%, accuracy = 91%). Automated classification of the apps' medical specialties resulted in 278 of 341 apps correctly classified (accuracy = 82%), whereas 63 apps were classified in a medical specialty different than the correct one. Of the 17 apps related to sensory systems health care, 15 (88%) were correctly classified (including eight related to audiology), whereas two apps were characterized as related to cardiology and neurology, specifically two apps for S&L due to use of terms such as stroke and brain injury in the app description (sensitivity = 88%, specificity = 95%, accuracy = 95%).

For a general picture of the performance of the automated algorithm, Figure 2b shows the percentage of apps correctly classified in each medical specialty. For nine of 19 medical specialties (dentistry, diabetes care, gastroenterology,

immunology and endocrinology, nutrition, oncology, pediatrics, sleep and respiratory care, and surgery), the automated algorithm was able to correctly characterize 100% of the apps. In five medical specialties (cardiology, fitness & wellness, gynecology and obstetrics, mental health and neurology, and sensory systems health care), the percentage was about 90%. In three specialties (dermatology, emergency medicine, and physiatry and orthopedics), the percentage was about 80%, whereas in urology and across specialties, the percentage of correctly classified apps was lower (67% and 53%, respectively).

Discussion

In this exploratory study, the basic modules of an automated method were developed to characterize the features of mobile apps by analyzing the information reported on the app store webpages. A sub–data set of 400 apps from the M and H&F categories on the U.S. iTunes app store was used to compare the classification relevant to apps' medical specialty performed by manual "gold standard" with the proposed automated method by extracting and analyzing UMLS terms and MeSH codes from the unstructured text description reported on the store.

Manual Characterization

- Medical specialties: Characterization of apps across medical specialties showed that several apps (about 15%) were not related to health or medicine, although they were labeled as M and/or H&F by the developer. Of the apps related to health and medicine, as expected, many were for Fitness & Wellness (about 40%). Interestingly, a significant percentage of the apps was not related to a singular medical specialty or condition, but to medicine or health care in general (across specialties, 22%).

- Promoters: The distribution of promoters in the testing set showed that companies and individual developers represent the lion's share of the market, in line with previous reports (Paglialonga, Riboldi, et al., 2017; Research2Guidance, 2017). However, the tested sample also showed the presence of apps promoted by stakeholders from the health care services area and from professional and scientific organizations. This fact can be considered as an encouraging trend in terms of expected quality and evidence-based research for apps, as medical professional involvement is one of the elements of evidence base, along with adherence to guidelines and recommendations, scientific validation, and reliability of app developers (Leigh, Ouyang, & Mimnagh, 2017; Paglialonga, Lugo, et al., 2018). Some apps in the set were found promoted by more than one category, for example, health care providers and professional & scientific associations, or health care providers and governmental institutions. This is in line with recent recommendations that highlight the need for multidisciplinary and multistakeholder

efforts in the development of health apps (Molina Recio, García-Hernández, Molina Luque, & Salas-Morera, 2016).

- Offered services: In the area of offered services, the high percentage of apps for tracking and practicing found in our testing set was likely due to the combined presence of many apps for fitness and lifestyle management (i.e., including tracking of activity, health data, daily habits, or rehabilitation programs). Overall, most of the apps offered tracking/practicing or education/learning functionalities (74% overall), whereas a relatively small number of apps provided telemedicine services, health record management, patient management services, or clinical support tools. These findings suggest that there is probably a room for further developments of apps in this field to support clinical service delivery and patient management. Further investigations are needed to understand the clinical needs and the potential benefits of apps in this area.

- Target users: In terms of target groups, the high number of apps for the general public (citizens: about 50%) was largely related to the fact that our testing set included many apps (61.2%) in the H&F category. Nevertheless, it is interesting to note that more than one of four apps (about 27%) were intended for use by patients and one of five (about 20%) were developed for professional use.

Overall, from the analysis of apps in the testing set, it seems that, for a high percentage of apps, the potential opportunities offered by mobile apps and smartphone capabilities are probably not fully exploited yet. In general, it appears that the offered services and functionalities are frequently the basic ones, especially tracking and delivery of information, whereas other services more related to the clinical workflow (health records, patient management, and telemedicine) are less represented. Moreover, from a technical point of view, currently available apps tend to behave more as basic software products than as fully capable, multisensor, connected apps. Future analysis of the promoters and offered services across the whole market of apps would be useful to highlight the actual trends and gaps in this area.

Automated Characterization

Automated characterization of apps in the testing set across medical specialties showed encouraging results. The first version of the method developed here provided 95% accuracy for the identification of apps related to sensory systems health care, 91% accuracy for the identification of apps related to health or medicine, and, on average, 82% accuracy for the classification of apps into medical specialties (range = 53%–100%). It is worth noting that these results are preliminary and that some of the percentage values in Figure 2b were computed from low numbers, as some medical specialties were represented by only a small sample of apps in the testing set (e.g., less than five apps in

gastroenterology, immunology and endocrinology, pediatrics, sleep and respiratory care, surgery, and urology). This led to a high variability of percentage estimates, so further verification over larger manually classified data sets is mandatory. Nevertheless, these preliminary results are useful as they suggest practical indications for improvement, especially for those medical specialties with lower classification accuracy, as outlined in the following paragraphs.

Improvements Needed

- Presence of general terms: A closer look at the mismatches between manual and automated characterizations showed that, when failures of the algorithm were observed, they were frequently related to the presence of general terms in the description. Therefore, MetaMap could hardly classify these apps as related to health or medicine. The observed trends suggest that the identification of apps that are across specialties (i.e., apps related to general medicine, medical education, nursing, or health care in general) poses major problems. This is likely due to the inherent tendency of MetaMap to characterize terms, including the most general ones, as related to specific topic areas (Aronson, 2001). For the same reason, our algorithm might tend to characterize as related to health or medicine some apps that have no medical content. In fact, specificity was found relatively low, that is, 56%, for the identification of apps related to health or medicine. Therefore, additional rules (e.g., introducing thresholds on general terms, computing additional scores, and creating more complex classification rules) and ad hoc vocabularies should be developed and included in future versions of the algorithm to better identify the medical content of apps and to correctly characterize those that have general medical content (i.e., across specialties).

- Partially overlapping categories: For some categories, classification mismatches were observed because they were inherently close to each other in terms of related vocabularies. For example, Fitness & Wellness shares several terms with Cardiology and Physiatry & Orthopedics. Similarly, Sleep and Respiratory Care has many terms in common with Mental Health & Neurology and Fitness & Wellness (especially related to relaxation and stress relief). Therefore, it would be important to further develop the algorithm and improve the rules for the identification of apps in these partially overlapping categories. This can be done, for example, by introducing additional vocabularies including common language terms to complement the highly specific UMLS terminology used by MetaMap. Extension of the algorithm to common language vocabularies is ongoing and will be tested in future studies.

- Improvements in sensory systems health care: The performance observed for sensory systems health care (sensitivity = 88%, specificity = 95%, accuracy = 95%)

was above the average performance of the algorithm, although there is a room for improvement in this medical specialty as well. In particular, for apps related to HHC, S&L, and audiology at large, it will be important to improve the associated vocabularies and to introduce more robust classification rules. For example, it would be important to introduce a double-check or an improved scoring system for general terms related to audiology (i.e., hear, speak, listen, and perceive) as these terms are indeed of very general use and, also, are frequently used in apps for mental health (e.g., to describe apps for peer-to-peer counseling, group therapy) or wellness (e.g., to describe apps for relaxation, meditation, and stress relief).

- Analysis of semantic types: In general, it might be useful to investigate if, in addition to the analysis of terms and scores, the analysis of semantic types (as retrieved by the MetaMap algorithm) might assist in a more robust characterization of medical specialties. For example, some semantic types (i.e., biomedical occupation or discipline, daily or recreational activity, and food) are likely to be revealing about the medical specialty and topic area and could thus be taken into account in the analysis. This could be done by introducing additional weighting factors or elements in the classification rules. Similarly, semantic types such as animal and functional concept can assist in the identification of apps not related to health or medicine, namely, recreational apps and games, respectively. In general, it will be important to further evaluate in future studies which semantic types to include and how to analyze them as well as the potential improvements in classification accuracy.

- Analysis of additional features: An extended future version of the method will be able to characterize also additional features of apps (promoters, offered services, and target users). As these features are not strictly related to medical concepts, different approaches than the one developed here will be used. A new version of the function "feature extractor" is currently under development and will be validated in future studies. It will make use of the IBM Watson Alchemy (IBM, Armonk, NY) text analytics tool in addition to Meta-Map, and it will include ad hoc lists of keywords and advanced classification algorithms to identify the different categories of promoters, services, and users, among those listed above and coded in previous studies (Paglialonga, Pinciroli, & Tognola, 2017; Paglialonga, Riboldi, et al., 2017). Upgrade of the algorithm by the development of functions for automated characterization of additional features is a necessary step toward the development of methods for automated characterization of apps along multiple components. In particular, in the area of HHC, it would be important to extract features of interest such as in the ALFA4-Hearing model, so as to characterize apps and compare them against a common structured schema. Extracting these features (e.g., the apps' promoters, offered

services, and target users) will be a key preliminary step to understand if, and to what extent, this kind of descriptive models can be filled out automatically, whenever new apps are launched on the market or periodically updated. Such an approach could lead to the development of near real-time descriptive app labels that could keep pace with the very rapid development of the market and that might summarize (or complement) the information provided on the app stores and developers' websites. In fact, it is acknowledged that the information reported on the vendor markets is frequently fragmented and sometimes incomplete or misleading as it is mostly at the discretion of the apps' developers (Paglialonga, Lugo, et al., 2018). Therefore, extraction of features from the apps' description reported on the app stores has an inherent limitation as these stores could provide only a partial picture of the various features of the apps. Nevertheless, the automated method proposed here can be, in principle, adapted for use in various platforms. As such, it could be used to extract meaningful information by combining several sources, including the ones more scientifically sound, such as app clearinghouses, expert review websites, or scientific/professional communities in the area of mHealth.

Conclusions

Although preliminary, the findings from this exploratory study suggested that automated methods based on text analytics, such as the one developed here, could be helpful to extract meaningful information about the medical specialty from the app stores' webpages. Such automated methods are able to rapidly analyze very large samples of apps' description and might assist in the selection based on a priori classification to answer specific needs. The ability to analyze a large database of apps along several features provides interesting opportunities to assess the distribution and pattern of features in the market or in medical specialties of interest. This could be useful to highlight possible trends and gaps in this emerging area, as well as challenges and opportunities for further development, and to monitor these trends along time. From a clinical point of view, the availability of automated tools for app filtering and characterization could be a valuable opportunity for health care professionals as it could support them in an informed, aware selection of health apps to be recommended to their patients.

Further developments are needed to improve classification performance, also including additional features and web sources and assessing the factors that influence classification accuracy, so as to take full advantage of the potential offered by natural language processing for the analysis of large databases. Compared with conventional keyword-based search, the proposed method can be the basis for a filtering tool that is context-aware as it is based on computational linguistic techniques and, also, it includes algorithms to estimate the probability of the correct interpretation of terms

and phrases (i.e., the MetaMap scores) as well as optimized rules that enable classification into multiple categories. Following improvement and optimization of the method, a quantitative analysis of its performance and computational running time compared with conventional keyword search will be essential to assess the benefit of this approach and its reliability for use in real-time characterization of apps.

Another strategic direction for improving the proposed method and to further support clinicians in their practice would be to explore ways to include direct or indirect measures of quality. This is particularly challenging as it is difficult to identify the core components of quality as well as appropriate measures to assess them. When dealing with mobile apps, quality includes, for example, components related to evidence base (adherence to guidelines and recommendations, scientific validation, reliability of app developers, and professional involvement in app development), trustworthiness (transparency, data management and data reuse, and data protection and privacy), and user-oriented quality (usability, quality of experience, functional design and aesthetics, ease of use, and perceived quality of content and function; Paglialonga, Lugo, et al., 2018). How to combine automated characterization of descriptive features with possibly automated characterization of quality measures is an entirely open question. It will be important in future studies to investigate whether some methods, among those proposed in the literature, could be used or adapted to be included in an automated approach. This would be of great value to fully empower clinicians with tools to assess and compare the quality of available apps with the ultimate goal of providing greater benefit to patients. Moreover, larger availability of comprehensive automated methods for app characterization and quality assessment, open to the community, may ultimately contribute to make developers more responsible and committed to deliver high-quality apps.

Acknowledgments

References

Aho, A. V., & Ullman, J. D. (1992). Patterns, automata, and regular expressions. In A. V. Aho & J. D. Ullman (Eds.), *Foundations of computer science* (pp. 529–590). Retrieved from http://infolab.stanford.edu/~ullman/focs.html

Aronson, A. R. (2001). Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. *Proceedings of the AMIA Symposium, 2001,* 17–21.

Aungst, T. D., Clauson, K. A., Misra, S., Lewis, T. L., & Husain, I. (2014). How to identify, assess and utilise mobile medical applications in clinical practice. *International Journal of Clinical Practice, 68,* 155–162. https://doi.org/10.1111/ijcp.12375

BinDhim, F. N., & Trevena, L. (2015). Health-related smartphone apps: Regulations, safety, privacy and quality. *BMJ Innovation, 1,* 43–45. https://doi.org/10.1136/bmjinnov-2014-000019

Boudreaux, E. D., Waring, M. E., Hayes, R. B., Sadasivam, R. S., Mullen, S., & Pagoto, S. (2014). Evaluating and selecting mobile health apps: Strategies for healthcare providers and healthcare organizations. *Translational Behavioral Medicine, 4,* 363–371. https://doi.org/10.1007/s13142-014-0293-9

Bright, T., & Pallawela, D. (2016). Validated smartphone-based apps for ear and hearing assessments: A review. *JMIR Rehabilitation and Assistive Technologies, 3*(2), e13. https://doi.org/10.2196/rehab.6074

European Union of Medical Specialists. (n.d.). *Medical specialties.* Retrieved from https://www.uems.eu/about-us/medical-specialties

Franko, O. I., & Tirrell, T. F. (2012). Smartphone app use among medical providers in ACGME training programs. *The Journal of Medical Systems, 36,* 3135–3139. https://doi.org/10.1007/s10916-011-9798-7

Grundy, Q. H., Wang, Z., & Bero, L. A. (2016). Challenges in assessing mobile health app quality: A systematic review of prevalent and innovative methods. *The American Journal of Preventive Medicine, 51,* 1051–1059. https://doi.org/10.1016/j.amepre.2016.07.009

HTML Working Group team at the World Wide Web Consortium. (2017). *HTML 5.2 W3C Recommendation, 14 December 2017.* Retrieved from https://www.w3.org/TR/html/

Jungnickel, T., von Jan, U., & Albrecht, U. V. (2015). AppFactLib—A concept for providing transparent information about health apps and medical apps. *Studies in Health Technology and Informatics, 213,* 201–204. https://doi.org/10.3233/978-1-61499-538-8-201

Kim, B. Y., & Lee, J. (2017). Smart devices for older adults managing chronic disease: A scoping review. *JMIR mHealth and uHealth, 5*(5), e69. https://doi.org/10.2196/mhealth.7141

Leigh, S., Ouyang, J., & Mimnagh, C. (2017). Effective? Engaging? Secure? Applying the ORCHA-24 framework to evaluate apps for chronic insomnia disorder. *Evidence-Based Mental Health, 20,* e20. https://doi.org/10.1136/eb-2017-102751

Misra, S., Lewis, T. L., & Aungst, T. D. (2013). 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 Dermatology, 149,* 661–662. https://doi.org/10.1001/jamadermatol.2013.606

Molina Recio, G., García-Hernández, L., Molina Luque, R., & Salas-Morera, L. (2016). The role of interdisciplinary research team in the impact of health apps in health and computer science publications: A systematic review. *BioMedical Engineering Online, 15,* 77. https://doi.org/10.1186/s12938-016-0185-y

Paglialonga, A., Lugo, A., & Santoro, E. (2018). An overview on the emerging area of identification, characterization, and assessment of health apps. *Journal of Biomedical Informatics, 83,* 97–102. https://doi.org/10.1016/j.jbi.2018.05.017

Paglialonga, A., Mastropietro, A., Scalco, E., & Rizzo, G. (in press). The mHealth. In P. Perego, G. Andreoni, & E. Frumento (Eds.), *mHealth between reality and future. EAI/Springer innovations in communication and computing.* Cham, Switzerland: Springer.

Paglialonga, A., Pinciroli, F., & Tognola, G. (2017). The ALFA4-Hearing Model (At-a-Glance Labelling for Features of Apps for Hearing Health Care) to characterize mobile apps for hearing health care. *American Journal of Audiology, 26*(3S), 408–425. https://doi.org/10.1044/2017_AJA-16-0132

Paglialonga, A., Pinciroli, F., & Tognola, G. (in press). Apps for hearing health care: Trends, challenges and potential opportunities.

In E. Saunders (Ed.), *Tele-audiology and the optimization of hearing health care delivery*. Hershey, PA: IGI Global.

Paglialonga, A., Pinciroli, F., Tognola, G., Barbieri, R., Caiani, E. G., & Riboldi, M. (2017). e-Health solutions for better care: Characterization of health apps to extract meaningful information and support users' choices. *2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI), 2017,* 1–6. https://doi.org/10.1109/RTSI.2017.8065915

Paglialonga, A., Riboldi, M., Tognola, G., & Caiani, E. G. (2017). Automated identification of health apps' medical specialties and promoters from the store webpages. *2017 E-Health and Bioengineering Conference (EHB), 2017,* 197–200. https://doi.org/10.1109/EHB.2017.7995395

Paglialonga, A., Tognola, G., & Pinciroli, F. (2015). Apps for hearing sciences and care. *American Journal of Audiology, 24,* 293–298. https://doi.org/10.1044/2015_AJA-14-0093

Payne, K. F., Wharrad, H., & Watts, K. (2012). Smartphone and medical related app use among medical students and junior doctors in the United Kingdom (UK): A regional survey. *BMC Medical Informatics and Decision Making, 12,* 121. https://doi.org/10.1186/1472-6947-12-121

Research2Guidance. (2017, November). *mHealth app economics 2017: Current status and future trends in mobile health.* Retrieved from https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health

Scherer, E. A., Ben-Zeev, D., Li, Z., & Kne, J. M. (2017). Analyzing mHealth engagement: Joint models for intensively collected user engagement data. *JMIR mHealth and uHealth, 5*(1), e1. https://doi.org/10.2196/mhealth.6474

Singh, K., Drouin, K., Newmark, L. P., Lee, J., Faxvaag, A., Rozenblum, R., . . . Bates, D. W. (2016). Many mobile health apps target high-need, high-cost populations, but gaps remain. *Health Affairs (Millwood), 35,* 2310–2318. https://doi.org/10.1377/hlthaff.2016.0578

Terry, K. (2015, June 22). Prescribing mobile apps: What to consider. *Medical Economics.* Retrieved from http://medicaleconomics.modernmedicine.com/medical-economics/news/prescribing-mobile-apps-what-consider?page=full

Wyatt, J. C., Thimbleby, H., Rastall, P., Hoogewerf, J., Wooldridge, D., & Williams, J. (2015). What makes a good clinical app? Introducing the RCP health informatics unit checklist. *Clinical Medicine (London), 15,* 519–521. https://doi.org/10.7861/clinmedicine.15-6-519

Yousuf Hussein, S., Swanepoel, D. W., Mahomed, F., & Biagio de Jager, L. (2018). Community-based hearing screening for young children using an mHealth service-delivery model. *Global Health Action, 11,* 1467077. https://doi.org/10.1080/16549716.2018.1467077

Zhang, Y., & Koch, S. (2015). Mobile health apps in Sweden: What do physicians recommend? *Studies in Health Technology and Informatics, 210,* 793–797. https://doi.org/10.3233/978-1-61499-512-8-793