



Article

WtsWrng Interim Comparative Effectiveness Evaluation and Description of the Challenges to Develop, Assess, and Introduce This Novel Digital Application in a Traditional Health System

Laura Sampietro-Colom ^{1,2,*} , Carla Fernandez-Barcelo ¹ , Ismail Abbas ¹ , Blanca Valdasquin ¹, Nicolau Rabasseda ², Borja García-Lorenzo ^{1,3} , Miquel Sanchez ⁴, Mireia Sans ^{5,6}, Noemi Garcia ⁵ and Alicia Granados ²

¹ Assessment of Innovations and New Technologies Unit, Research and Innovation Directorate, Clínic Barcelona University Hospital, 08036 Barcelona, Spain

² Mangrana Ventures S.L., 08006 Barcelona, Spain

³ Kronikune Institute for Health Sciences Research, 48902 Barakaldo, Spain

⁴ Emergency Department, Clínic Barcelona University Hospital, 08036 Barcelona, Spain

⁵ CAP Comte Borrell, Consorci Atenció Primària Salut Barcelona Esquerra—CAPSBE, 08029 Barcelona, Spain

⁶ Health 2.0 Section of the Col·legi Oficial de Metges de Barcelona, 08017 Barcelona, Spain

* Correspondence: lsampiet@clinic.cat



Citation: Sampietro-Colom, L.; Fernandez-Barcelo, C.; Abbas, I.; Valdasquin, B.; Rabasseda, N.; García-Lorenzo, B.; Sanchez, M.; Sans, M.; Garcia, N.; Granados, A. WtsWrng Interim Comparative Effectiveness Evaluation and Description of the Challenges to Develop, Assess, and Introduce This Novel Digital Application in a Traditional Health System. *Int. J. Environ. Res. Public Health* **2022**, *19*, 13873. <https://doi.org/10.3390/ijerph192113873>

Academic Editors: Devidas Menon and Tania Stafinski

Received: 12 September 2022

Accepted: 22 October 2022

Published: 25 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: Science and technology have evolved quickly during the two decades of the 21st century, but healthcare systems are grounded in last century's structure and processes. Changes in the way health care is provided are demanded; digital transformation is a key driver making healthcare systems more accessible, agile, efficient, and citizen-centered. Nevertheless, the way healthcare systems function challenges the development (Innovation + Development and regulatory requirements), assessment (methodological guidance weaknesses), and adoption of digital applications (DAs). WtsWrng (WW), an innovative DA which uses images to interact with citizens for symptom triage and monitoring, is used as an example to show the challenges faced in its development and clinical validation and how these are being overcome. To prove WW's value from inception, novel approaches for evidence generation that allows for an agile and patient-centered development have been applied. Early scientific advice from NICE (UK) was sought for study design, an iterative development and interim analysis was performed, and different statistical parameters (Kappa, B statistic) were explored to face development and assessment challenges. WW triage accuracy at cutoff time ranged from 0.62 to 0.94 for the most frequent symptoms attending the Emergency Department (ED), with the observed concordance for the 12 most frequent diagnostics at hospital discharge fluctuating between 0.4 to 0.97; 8 of the diagnostics had a concordance greater than 0.8. This experience should provoke reflective thinking for DA developers, digital health scientists, regulators, health technology assessors, and payers.

Keywords: mHealth; innovation; technology assessment; digital health; digital application; triage; WtsWrng



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Context Analysis

Access to essential healthcare services represents a core challenge, politically and professionally, for both developing and developed countries [1–3]. According to the World Health Organization in December 2017, 50% of the world's population lacks access to essential healthcare services [2]. Therefore, there is *limited access* to appropriate basic health counseling for the most frequent, common symptoms, which is an ideal time for making the decision regarding the need and type of health care. In developed countries, there is an

inappropriate use of healthcare resources, which may lead to further difficulties accessing services for those most in need [4–6]. In the USA, several reasons for inappropriate use have been identified, including the lack of a proper primary healthcare system that responds to basic health concerns and access difficulties to the appropriate level of care due to multiple structural, financial, and logistical barriers [6]. Additionally, healthcare executives mention that the inappropriate use of the emergency department (ED) for low-severity issues is one of the most critical problems EDs face nowadays [7]. Inappropriate use of health resources is an opportunity-cost loss, as attending low-severity cases at EDs precludes using human knowledge and expertise in other more severe and complex cases. Moreover, the COVID-19 pandemic has shown an increased *demand in ED visits* [7], adding to the already difficult circumstances given the trend towards a decreasing workforce due to shortage or burnout [7]. A shortage of doctors and nurses is now a global challenge, with an expected deficit of 18 million by 2030 [8]. Further access challenges to basic healthcare services are seen in lengthy waiting times and patient–physician miscommunication [9]. *Waiting times at ED* are one of the most frequent complaints registered by users [7,9]. Scientific societies for emergency medicine have set up appropriate waiting times according to the level of severity of the problem (ranging from immediate attention to several levels of “reasonable” waiting times according to risks and acuity) [10–12]. However, these times are not always kept [10]. Another issue is *patient–physician miscommunication* which may lead to suboptimal care (including harm and death), decreased understanding of the diagnosis and disease management, and difficulties in shared decision-making [13,14]. Some causes of miscommunication include age-related challenges, reduced capacity or disability, poor language, communication skills, cultural barriers, and literacy and health literacy [13,14]. Therefore, clarity in the written and spoken language at medical encounters is warranted and requested [1].

Besides the current weaknesses related to appropriate care access, traditional health systems are challenged by a shift in attitudes and behaviors of citizens towards greater *consumer engagement and empowerment* in healthcare decisions. A survey on consumer priorities in health care showed that consumers want access to care when, where, and how best suits them; to comply with these expectations, they are using technology [15], e.g., self-diagnosis with search engines. A survey administered to 1013 United Kingdom (UK) adults revealed that Google was the first point of reference for people as a primary go-to tool when feeling unwell or experiencing potential illness symptoms. Importantly, the core reason for doing so appears to be in response to the difficulty in making a GP appointment [16,17].

All the aforementioned challenges for current traditional health systems, along with the changing trends of society, are pushing health systems to rethink the way health care is organized and provided. New care models are required to ensure quality and access to health care for patients. Furthermore, and to achieve this, costs, behavior changes, and technological progress must be effectively managed. A transformational change in the way health care is organized and provided through a shift towards a patient-centered healthcare paradigm is one of the best options [17], and as seen during the pandemic, emerging digital health technologies have the potential to drive forward said change [18,19]. Digital health technologies are perceived as the tool helping to face demographic challenges (and their associated increasing care demand), to maintain health systems as economically sustainable, and to satisfy citizen expectations [17,20,21]. Moreover, they are being proposed (and starting to be used) to improve access to healthcare services [22] and, for some, decrease miscommunication problems [14].

Web-based tools and digital applications (DAs) for counseling when a symptom appears have been developed as first aid for patients and citizens [22]. The idea behind these tools is both to empower citizens in the self-management of their health and to properly drive potential users to healthcare resources when needed (and discourage the access of health care if there is no need). These tools are used as a triage tool (advising the citizen of a course of action: self-care, making a general practitioner appointment, or

going to an emergency department) or as a symptom checker (which gives a presumptive diagnosis to the user based on their reported symptom/s) [22–25].

Though efforts have been made to introduce digital technologies (including DAs) in healthcare systems, resistance in uptaking innovation for breakthrough medical technologies has been seen in the past [26]. Likewise, digital technologies have been developed and updated constantly, which leads to a development uptake gap as healthcare systems are unable to implement these new innovations as soon as they are released. Moreover, most of these innovative DAs are developed by start-ups that face the challenges of introducing technologies aimed at changing the healthcare “status quo” under Innovation + Development, regulatory, assessment, and healthcare adoption requirements designed for very different types of technologies (i.e., medicines, medical devices, IV diagnostics) and for an “outdated” healthcare paradigm (i.e., mostly in person, mostly non-patient empowerment), which hinders their adoption.

For any innovation, the path to develop and move from a conceptual idea to a real product to the market follows specific steps well known in the process of management of innovation, including development of the product, intellectual protection, proof-of-concept and clinical validation, regulatory approval (CE mark in Europe), healthcare system coverage/reimbursement, and healthcare adoption based on the value the new technology brings. For each step, different agents act in deciding the go/no go for the technology. The recently issued EU Medical Device Regulation (MDR) [27] considers digital health technologies, including DAs, as a medical device (MD). Under this regulation, the Notified Bodies (NB) are the institutions that, upon reviewing the results from clinical studies and other technical information, grant access to the European market (CE mark). The CE mark is necessary but not always enough to obtain health authorities’ reimbursement in Europe. Health Technology Assessment (HTA) institutions assess the value of innovative technologies for a specific healthcare system using a predetermined set of criteria and methods, producing recommendations for decisions on coverage/reimbursement by healthcare authorities [28,29]. Finally, adoption of an innovation by the healthcare provider will depend on a myriad of factors (e.g., how the innovation is going to be financed, acceptance by users). Trying to prove value and reach the market for some DAs is filled with obstacles and challenges, especially for start-ups and small and medium enterprises (SME). This process becomes more challenging when regulatory and healthcare systems are still learning how to manage specific disruptive innovations.

The objective of this paper is to present the development and interim comparative effectiveness evaluation of a novel DA called *WtsWrng* (WW), according to the current requirements for new technologies to access the European market. In doing so, we also aimed to show the challenges faced and the novel methodological approaches used to evaluate WW.

1.2. WtsWrng (WW) DA: Why and for What Purpose

In the initial phases of developing WW, the first step was the identification of the unmet need, the target population, and where the technology will be used in the pathway of care. From the beginning, it was clear that WW aimed to be a tool to offer citizens quicker and more agile access to answers to their basic health concerns and to be placed as an additional virtual access option for current in-person health care. It was aimed to empower citizens in their decisions about how to proceed when a symptom appears. Though few similar online and smartphones DAs already exist [22], weak scientific evidence of their effectiveness is available [30]. Additionally, all available DAs are built using written and oral language systems, which are inherently subject to miscommunication in the patient–physician encounter, often leading to medical errors as mentioned above [1,13,14]. Moreover, using written and spoken language also limits the worldwide scalability of the product, even when using artificial intelligence (AI)-based translators. WW creators developed a solution to overcome these challenges by both designing a DA using the universal language of images to interact with citizens and aiming to generate strong

evidence of its effectiveness comparing WW with the current standard of care at the ED and primary care center (PCC).

The WW DA is a digital triage, symptom checking, and monitoring solution that can be downloaded on any smartphone. Given a reported symptom, WW provides suggestion on basic, non-prescriptive behavior using only images, intentionally avoiding the use of written or voice language. This unique feature makes WW a disruptive DA since, as of today, we are not aware of any triage DA that interacts with users via images only. The user navigates through a series of encoded drawings regarding his/her symptoms and, after a reduced digital anamnesis (maximum 2 min), suggests a course of action (self-care, appointment with general practitioner, specialist, emergency department, or immediate help is needed). Figure 1 shows the patient journey of WW compared with the current access to care.

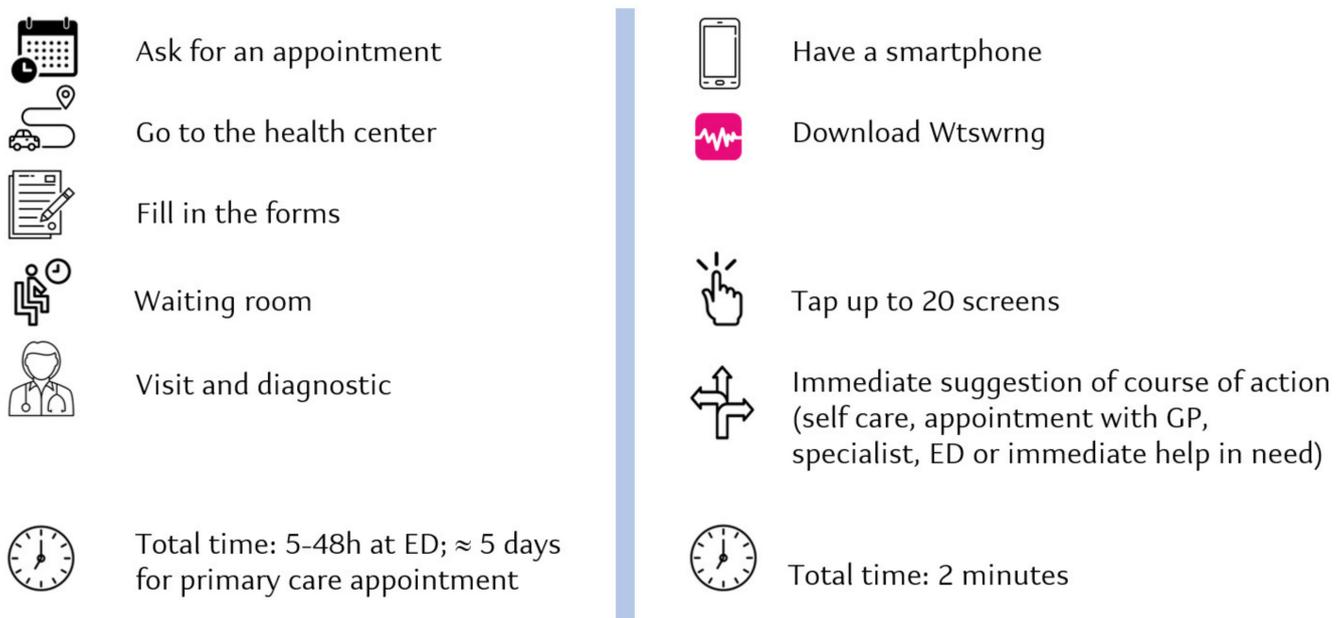


Figure 1. Patient journey: conventional care versus WtsWrng. ED: Emergency Department; GP: General Practitioner.

WW displays the 20 most frequent symptoms that lead to an ED or a non-scheduled PCC visit on the user’s smartphone [16,31,32]. The user selects the symptom and scrolls through screens, where different images appear, reflecting aspects of usual anamnesis performed by clinicians in person to help patients in thinking about a potential presumptive diagnosis linked to their initial symptom/s. With the time reduced, digital anamnesis leads to a set of presumptive differential diagnoses. These presumptive differential diagnostics are kept in the background and are not visible to the user. They are used to stratify users by risk of severity and, consequently, suggest the most appropriate course of action. Figure 2 shows the WW interface.

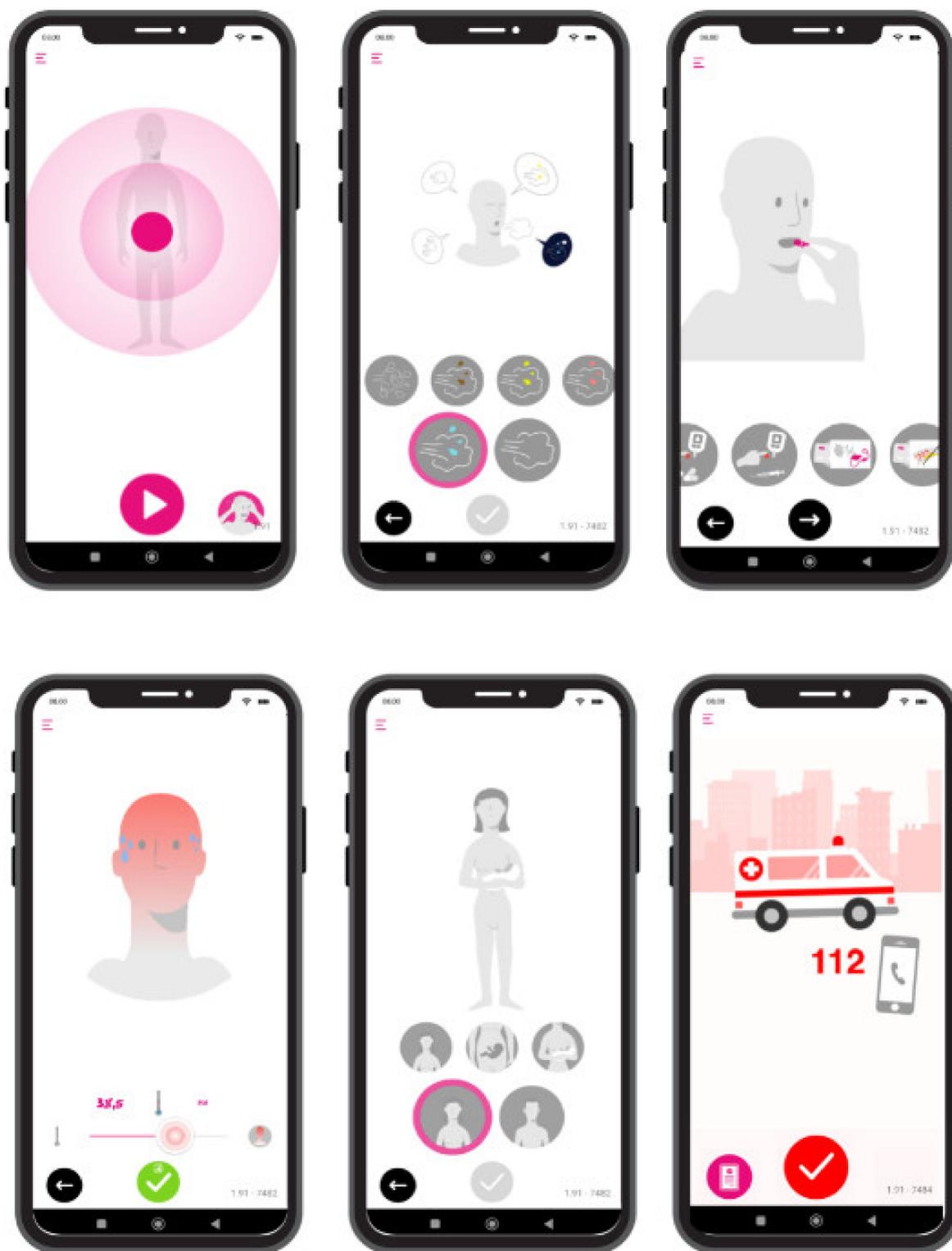


Figure 2. Sample of images shown on WW screens to interact with citizens.

2. Methods: WtsWrng Development and Evaluation

2.1. Building the First Prototype: Ensuring the Accuracy of WtsWrng Decision Trees and the Subsequent Algorithms and Drawing Design

A total of 20 symptoms were identified as the most frequent reasons for attending the ED and non-scheduled PCC visits. These symptoms come from a mix of different sources: the results of a field study performed at a hospital (Clinic Barcelona Univer-

sity Hospital—CB) and PCC in our country (Catalonia/Spain), from selected literature reviews [16,31–33] (to ensure transferability), and from an expert panel consisting of senior emergency physicians that discussed the symptoms to be included. The symptoms digital anamnesis led to 336 presumptive differential diagnoses, including diagnosis for some treatable rare diseases.

A multidisciplinary team reviewed the medical literature, looking for evidence-based symptoms guidance that is used in primary care and emergency room triages [34–40], and these evidence-based resources were used to develop the core decision trees (i.e., from each of the 20 symptoms triggering a visit to doctor, to the 336 presumptive differential diagnoses) [34–40]. Decision trees were then tested using clinical simulation in more than 6000 “near-live” clinical cases, as suggested by some authors [20]. The result of this technical validation was the basis to create the training rules to design a predictive machine learning-authored algorithm, with the aim of reducing the number of questions required until reaching the final WW suggestion on the course of action, as the number of users grow, to a similar number to that used during a standard/conventional presential or phone clinical anamnesis.

As previously mentioned, WW only uses images. To create the images, a review of international comics (from different continents) to select images of signs (e.g., pain) that could be clearly understood by different cultures was undertaken. The understanding of a first version of images was tested by a small convenience sample of people (colleagues, friends, family) from different cultures (Caucasian, Hispanic, black people). To date, WW includes more than 5000 encoded original and protected drawings.

2.2. Usability and User Experience

A small-scale usability (i.e., observing how users interact with the product) and user experience (UX) test was conducted to identify usability issues, and to assess use and acceptance. The test was performed on a sample of 100 citizens attending the ED of a hospital (n = 50) and its linked PCC (n = 50). The test was performed in-person by a field researcher. A questionnaire was designed and used to capture demographic characteristics (country, native language, gender, age), digital literacy (literacy, use and type of digital tool-smartphone, computer, tablet, smart watch, wearable, none; level of DA use, online buying, social network use), and UX (WW usefulness, intuitiveness, design liking, understanding of images, visibility of images, easiness of images selection, future use of WW, and recommendation of WW to others).

2.3. Clinical Validation of WtsWrng: From Fast-Track Early Scientific Advice from NICE on WW Evidence Generation Plan to the Study Design, Agile Iterative Development, and Agile Iterative Interim Analysis

A cross-sectional, single-arm clinical trial was designed to test the accuracy and precision of WW [41,42]. The study design aimed to demonstrate that WW (i.e., the intervention) is non-inferior (i.e., at least equal or better) in precision and accuracy than the tools currently used by a hospital ED to triage (i.e., Structural Emergency Triage. Andorra Triage Model-MAT) and to diagnose patients (i.e., the diagnosis written on the ED clinical record at patient discharge). These two comparators are considered the “gold standard” in this study. The study protocol was submitted to NICE (National Institute for Health and Clinical Excellence, UK) to obtain early scientific advice, under the fast-track review guidance for COVID-19 studies (request SA593) [43], leading to the inclusion of some of the advice received in the protocol. The final clinical protocol obtained the approval of the Ethics Committee of the Clinic Barcelona University Hospital (CB), where the evaluation was performed (Reg. CB/2020/0087).

WW was tested on a sample of consecutive patients attending the ED of CB from October 2020 to July 2021, having met the predefined inclusion criteria (i.e., patients over 18 years old, triaged to ED levels III, IV, or V, and conscious in order to use WW with a smartphone). Patients triaged to levels I and II were excluded from the study since

they were unconscious or very severely ill or injured in need of immediate attention and, therefore, are unable to properly use WW.

A field researcher attended the ED of the hospital and asked for voluntary participation. The patients were given a smartphone with WW and were asked to use it. Final suggestions for the user course of action were masked at this point (i.e., no suggestion on course of action was displayed on the screen as the end result). Written informed consent was obtained from all participants.

The triage system of the CB's ED uses a nurse-driven computer system (MAT [44]) that sends patients to different levels of care depending on five degrees of severity, as defined internationally [45]. To assess triage accuracy and the precision of WW with the ED system at the hospital, based on clinicians' suggestions, we considered that level III corresponds to the suggested course of action to go to hospital/specialized care, level IV corresponds to going to primary care (i.e., non-emergency cases that do not need to go to a hospital ED but to a PCC), and level V corresponds to self-managing the symptoms. Given a symptom, triage accuracy was calculated by means of comparing the percentage of patients correctly triaged by WW, given the results from the ED triage system. Additionally, to test triage precision, the percentage of patients correctly triaged by WW was compared to the results from the level assigned (i.e., III–V) by the ED triage system at the CB, by calculating the observed concordance, Kappa, and B statistics.

Precision and accuracy between the presumptive diagnosis given by WW and the clinical record diagnosis at patient discharge were also analyzed. As first choice, the ICD-10 coded diagnostic at discharge from the ED was extracted from the administrative database of the hospital designed for billing purposes. Their ICD-10 codes were considered the "gold standard to compare", with presumptive diagnosis given by WW (also coded with ICD-10). During the interim analysis, inaccuracies in the administrative database were detected; therefore, the clinical records at discharge were also reviewed to have the final diagnosis given to the patient for comparison. The diagnostic precision of WW was then studied by means of the concordance between the diagnostic of a patient at ED discharge (from clinical records) and the first 5 presumptive diagnoses provided by WW. The observed concordance, Kappa, and B statistic were applied to the 12 most frequent diagnostics (defined as $n \geq 9$ patients) at hospital discharge, to observe potential differences between the results when applying these statistical parameters. Kappa is the most used parametric statistic to assess concordance between a diagnostic health technology and the gold standard [46]. The estimated Kappa is a variable that follows a normal distribution with mean k in formula Figure 3a and standard deviation $S(k)$ in formula Figure 3b. To properly use Kappa as a value statistic, it requires both a high and similar number of observations in n_{11} and n_{22} , as well as similar numbers in n_{21} and n_{12} (see Table 1). The B statistic in formula Figure 3c is a non-parametric statistic that can also be used to assess concordance between a diagnostic health technology and the gold standard; it has an advantage over Kappa in that similar numbers are not required in n_{11} , n_{21} , n_{12} , and n_{22} [47], as observed in our study. Figure 3 shows the Kappa and Beta statistic calculations, while Table 1 shows the theoretical distribution of subjects by diagnostic sources (WW versus ED clinical record at discharge) and response category.

Table 1. Theoretical distribution of subjects by diagnostic sources and response category.

		WtsWrng		Total
		NO	YES	
Hospital	NO	n_{11}	n_{12}	g_1
	YES	n_{21}	n_{22}	g_2
	Total	f_1	f_2	

$$\hat{k} = \frac{p_o - p_e}{1 - p_e} \quad (\text{a}) \quad S(k) = \sqrt{\frac{p_o(1-p_o)}{n(1-p_e)^2}} \quad (\text{b})$$

$$\hat{B} = \frac{\sum_{i=1}^q n_{ii}^2}{\sum_{i=1}^q g_i f_i} = \frac{n_{11}^2 + n_{22}^2}{g_1 f_1 + g_2 f_2} \quad (\text{c})$$

Figure 3. Kappa and B statistic calculation formulas. Legend: where n is the total number of observations, n_{ii} is the n_{11} and n_{22} representing the number of agreements between hospital and WtsWrng; g_i is the g_1 and g_2 representing the total diagnoses identified and not identified by the hospital; f_i is the f_1 and f_2 representing the total diagnoses identified and not identified by WtsWrng; p_o is the proportion of the observed agreements between the hospital and WtsWrng $(n_{11} + n_{22})/n$; and p_e is the expected agreements $(g_1 g_2 + g_2 f_2)/n^2$ assuming independence between the assessment of hospital and WtsWrng.

Finally, presumptive diagnosis sensitivity and specificity of WW compared with the diagnostic at hospital discharge were calculated for the 12 most frequent diagnostics attending the ED.

At the time of clinical testing, WW was in a development stage (Technology Readiness Level 6) [48]. This meant that an agile iterative analysis was necessary, where quick rounds of testing–re-designing–testing were conducted [49]. It is also important to mention that the study was performed during the COVID-19 pandemic; therefore, evolving symptoms from this disease were also introduced into the decision trees that fed the algorithms during the study period.

Stata 13 software [50] was used for calculation purposes. Nevertheless, the B statistic was hand-calculated using specific formulas since Stata does not include a program for this purpose.

2.4. Protection of WW

The European Union granted WW access to IPA4SME, a program devoted to innovative European SMEs, to analyze ways to protect any asset [51]. Following the results from the audit, different protection modalities were performed.

3. Results

3.1. Usability and User Experience

Of the 100 users, 60% of the sample were women. Age ranged from 18–29 (15%) to more than 71 years old (21%, including a 94-year-old). Participants born in Spain comprised 72% of the sample, with the rest (28%) non-Spaniards (Lebanon, Pakistan, Sweden, and Latin-American countries). Native language included Spanish ($n = 52$), Catalan ($n = 38$), and a variety of other languages (Galician, Italian, Urdu, Portuguese, Swedish, English). One person was not literate. Smartphone users comprised 99% of the sample, while one person did not use any type of digital device. Of the 100 people using smartphones, 80 used DAs in their smartphone, 79 of 100 people they buy things online, and 60 of the sample used social networks. Based on a Likert Scale from 1–10 (1 not useful at all, 10 very useful), 81 people (79%) scored WW over 7, which shows that WW is perceived as a useful tool. They also reported a good understanding (60%), visibility (85%), and easiness in selecting the images on the smartphone (87%). Moreover, 90 participants (88%) liked the design (images) and 83 (81%) mentioned their willingness to use WW when available. Lastly, 87 people (85%) mentioned that they would recommend WW to others.

3.2. Clinical Validation: Agile Iterative Interim Analysis

A total of 567 patients who met the inclusion criteria were enrolled in the study. There were 117 patients excluded due to failures in the transmission of data to the website (transmission failures happened at different days and times during the study period). Of the remaining 450 patients, 45 did not have a diagnosis at discharge in their clinical record

(i.e., they usually had the same symptoms that originated the visit to the hospital ED) and were excluded from the analysis. Finally, from the 20 more frequent symptoms that trigger doctor visits, those symptoms with fewer than 10 patients were excluded from the analysis, accounting for a further 23 patients; the exclusion of these patients was made to avoid infra or overestimation of WW effectiveness. The data of the remaining 382 patients were finally included for the assessment of WW's precision and accuracy of triage and presumptive diagnosis. Table 2 shows the demographic characteristics of the 382 participants in the clinical study.

Table 2. Characteristics of sample (N = 382).

	n (%)
Gender	
Female	179 (47)
Male	203 (53)
Age	
18–29	43 (11)
30–39	37 (10)
40–49	35 (9)
50–59	67 (18)
60–69	78 (20)
70–79	80 (21)
80–99	42 (11)
Nationality	
Spain	312 (82)
Other European countries	14 (4)
Latin American countries	47 (12)
Others	9 (2)
Employment status	
Housekeeper	2 (1)
Self-employed	6 (2)
Student	13 (3)
Employed	128 (34)
Unemployed	37 (10)
Sick leave	19 (5)
Retired	177 (46)
Higher education	
Without studies	9 (2)
Primary school	73 (19)
Secondary school	34 (9)
High school	135 (35)
Bachelor's degree	103 (27)
Master's degree	18 (5)
PhD	10 (3)

Table 2. Cont.

	n (%)
User of technologies (Mobile, Tablet, etc.)	
YES	360 (94)
NO	22 (6)
Use of DA	
YES	248 (65)
NO	134 (35)
Use of health DA	
YES	165 (43)
NO	217 (57)

A WW triage accuracy ranging from 0.62 to 0.94 for the most frequent symptoms attending ED was observed. Figure 4 shows the results of triage accuracy by symptom. Triage precision showed an observed concordance (p_o) of 0.78, a Kappa of 0.39, and B statistic of 0.81. The B statistic was shown to be closer to the observed concordance than Kappa.

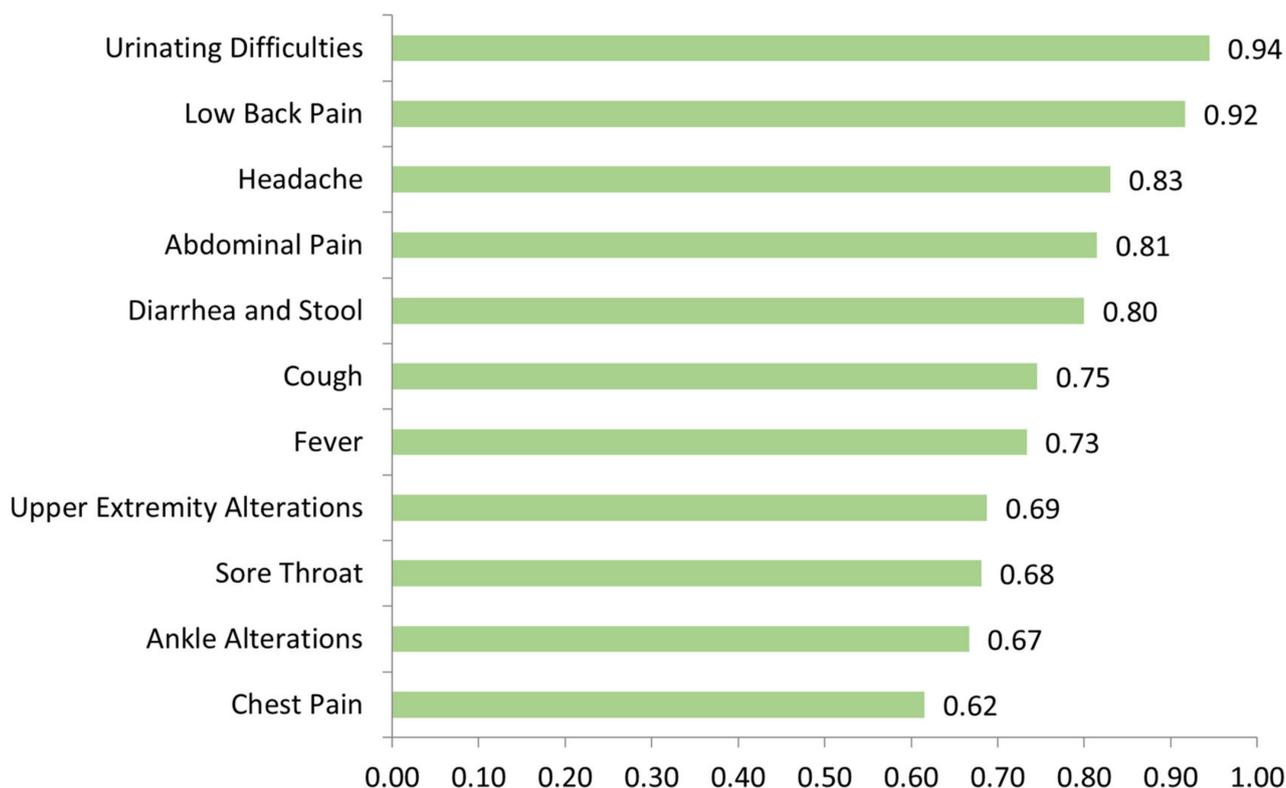


Figure 4. WW triage accuracy (from 0 to 1, with 1 equal to 100% accuracy) for the most frequent symptoms attending ED.

WW precision in presumptive diagnosis for the 12 most frequent diagnoses across the studied patients by means of observed concordance, Kappa, and B statistics is shown in Table 3.

The observed concordance (p_o) for the 12 most frequent diagnostics attending hospital ED fluctuated between 0.4 to 0.97; for 8 of the 12 diagnostics, the observed concordance was over 0.8. When using Kappa, the results ranged from 0.005 to 0.37, while the B statistic

ranged from 0.36 to 0.97, which is closer to the observed concordance (p_o). The B statistic continued to be closer to the observed coefficient of concordance than Kappa statistics. The sensitivity and specificity of WW were found to exceed 0.8 for 3 and 8 of the 12 most frequent diagnostics, respectively. Table 4 shows a summary of the results.

Table 3. Observed concordance, Kappa, and B statistics for the 12 most frequent diagnostics at hospital discharge (ordered for B statistics).

	n_{11}	n_{12}	n_{21}	n_{22}	g_1	f_1	g_2	f_2	p_o	k	B	Sens.	Spec.
Chronic kidney diseases	371	10	1	0	381	1	372	10	0.97	0.00	0.97	0.000	0.997
Lung cancer	368	5	7	2	373	9	375	7	0.97	0.23	0.97	0.286	0.981
Cellulitis	366	3	10	3	369	13	376	6	0.97	0.30	0.97	0.500	0.973
Gastroenteritis	354	5	16	7	359	23	370	12	0.95	0.37	0.94	0.583	0.957
Heart failure	343	21	16	2	364	18	359	23	0.9	0.05	0.90	0.087	0.955
Urinary infection	332	3	40	7	372	10	335	47	0.89	0.21	0.88	0.700	0.892
Asthma	333	4	42	3	337	45	375	7	0.88	0.09	0.88	0.429	0.888
Low back pain	309	2	66	5	311	71	375	7	0.82	0.10	0.82	0.714	0.824
Renal colic	292	3	80	7	295	87	372	10	0.78	0.10	0.77	0.700	0.785
COVID-19	136	10	162	74	146	236	298	84	0.55	0.20	0.38	0.881	0.456
Other respiratory disorders	144	2	227	9	146	236	371	11	0.4	0.02	0.37	0.818	0.388
COPD	141	1	225	15	142	240	366	16	0.41	0.04	0.36	0.938	0.385

Sens: Sensitivity; Spec: Specificity.

Table 4. WW precision and accuracy comparing the first 5 WtsWrng presumptive diagnoses and the diagnosis in ED clinical records at discharge.

Results (N = 382)	
Observed concordance (p_o)	0.4–0.97 (0.8 in 8 diagnostics) *
Kappa	–0.005–0.37
B statistic	0.36–0.97 (0.8 in 8 diagnostics) *
Sensitivity	0.09–0.94 (0.8 in 3 diagnostics) *
Specificity	0.39–0.98 (0.8 in 8 diagnostics) *

* Numbers in brackets: number of diagnostics, over the 12 most frequent diagnostics at hospital discharge, with a result ≥ 0.8 .

3.3. Protection of WW

Following the recommendations given by the IPA4SME program regarding intellectual protection modalities, WW was registered as a European Union Trademark/OHIM (#016480006). Main images were protected, while trade secret and copyright under Directive 2009/24/EC were also used as other barriers to disclosure. WW was also registered at the Benelux Office for Intellectual Property (iDEPOT # 113467).

4. Discussion

WW is the result of a compilation of digitalized evidence-based clinical decision trees that use supervised (i.e., as input data are fed into the algorithms, healthcare professionals check the output, with the weights of the algorithms adjusted until the model has been fitted appropriately; this is a cross-validation process activity) machine learning algorithms [52] to provide a disruptive (i.e., image-driven) clinical anamnesis and symptom triage to aid citizens in their decision regarding the need to see a healthcare professional, when and where, and to monitor any changes in the course of the suspected pathology, redirecting them to the appropriate level of care. WW aims to disrupt [53] and to be used as the

first resource globally for people that have a health concern (symptom), by using their smartphone. WW's quick and easy access can contribute to overcoming some of the current challenges that the traditional health system faces in line with some suggestions of systemic transformation for health systems [26]. Moreover, the new approaches proposed to develop and evaluate WW aims to inspire health stakeholders towards exploring new regulation requirements and policies for this type of technology.

Access to basic healthcare counseling, due to a lack of infrastructure [2,6] or long waiting times when an acute health problem appears [7], is a structural problem of current health systems. WW is easily accessible for smartphone users (83.40% of the world's population owns a smartphone [54]) and provides a suggestion of action in a short space of time (2 min), which can contribute to overcoming accessibility problems [54]. Importantly, although WW is based on authored algorithms, an API (Application Programming Interface) web can be developed to allow integration and compatibility with health system requirements, minimizing the technology country's adoption barriers. The results of our usability and UX initial study show a good acceptance of WW with a high willingness to use WW when available (81% of surveyed people) and disposition to recommend it to others (85% of surveyed people).

Inappropriate use of healthcare resources is also a challenge faced by health systems. In the UK, 61% of the visits at a hospital ED and PCC were not necessary [4], which represents an avoidable expenditure of EUR 900 million in 2020 [4]. In the USA, between 13.7% and 27.1% of visits to a hospital ED and PCC are not necessary, accounting for unnecessary expenditure of USD 4.4 billion yearly [5]. The results from the WW's interim analysis in a real-life clinical setting showed a good global precision as a triage tool (B statistic of 0.8, global observed concordance of 0.78). These results can position WW as a tool to guide a more appropriate citizen decision for using healthcare resources and, therefore, with the potential to avoid unnecessary expenditure. However, this claim may be disputed as some authors mention that the current triage DAs do not comply with their claim to reduce demand [55]. Conversely, others mention that it is not clear if the increase in demand of services observed may be due to the emergence of an unmet need or a supply-induced demand [56]. A specific study to test the impact of WW in decreasing demand and improving efficiency in healthcare resource use in a large real-life study, as suggested by some authors [57], is already planned.

A lack of a proper primary healthcare system that answers basic health concerns has also been identified in the USA as a reason for inappropriate ED use [6]. In this country, a review of 115,081 medical records showed that the avoidable top five ED disorders were mild complaints that a PCC can assess as a first contact [17]. In our study, WW has demonstrated good precision in identifying those patients that would have been more appropriately directed to primary care (0.95 observed concordance and 0.94 B statistic when comparing a hospital triage system—those patients directed to level IV—with WW results suggesting going to a primary care physician). Finally, research suggests that current DAs that are effective in triaging are not the same as those that have been shown to be good at diagnosing [55]. In this study, WW showed good specificity (over 0.8) for 8 of the 12 most frequent diagnoses attending ED, which compared favorably with similar DAs [58]. These results show the good performance of WW as both a triage and a presumptive diagnosis tool.

WW operates using images in its interaction with the user, aiming to both overcome the miscommunication problems linked to the current spoken and written medical encounters with patients and answering a rising social trend. Regarding the former, creative solutions are demanded [1]. As for the latter, images are identified as the tools to be used in future communication, including medicine. A current global trend is the rising use of digital tools using images (e.g., video games, where the number of users is expected to reach 2829.8 million by 2027) [59,60]. In the medical field, some professionals and authors are already claiming that they can facilitate communication in the medical field because of the avoidance of the ambiguity of words [61,62]. The usability and UX test performed for WW

in a sample of multicultural patients, with a different level of literacy, in a real-life clinical setting showed good visibility and understanding of images, which were also easy to select on the screen. Nevertheless, following users' suggestions, some images were improved at this point (more than 44 improvements were made).

Besides the societal trend to use more and more images in communication, there is also a trend towards both citizen empowerment in healthcare decisions [63] and immediateness in receiving an answer to basic health concerns, skipping an in-person visit that may not be necessary and takes time. As a digital health tool, WW overcomes the in-person, non-patient-centered approach to healthcare provision prevalent in traditional health systems [63]. Immediateness is answered nowadays by using digital tools. Currently, 6 out of 10 Europeans citizens go online when looking for health information; health care is the third-largest web activity across all generations [17]. According to the 2014 edition of the McKinsey report (N = 3000 consumers surveyed), for more than 90 respondents, DAs and websites are more effective ways to perform quick consultation activities than in-person or even phone consultations [63]. Another study showed that 59% of US adults have looked online for health information to make a decision or course of action; after digitally consulting, 38% thought that it was something they could take care of at home and 46% thought that they needed attention from a medical professional, while in 11% of cases, it was both or in-between [16].

Any related health DA devoted to suggesting, informing, or driving a course of action will have health consequences if the users act according to the information provided. Therefore, the credibility of the information used by any DA is critical [64]. This means that the background where the tool runs, as well as its clinical assessments, must be effective and of the highest-possible quality. In the case of WW, evidence (evidence-based guidance), experience (participation of healthcare professionals), and technology (supervised machine learning) were all considered in its development, ensuring minimal deviation from proper triage and diagnosis.

One of the evaluation activities needed to advance any emerging DA, before its market release and to promote user uptake, is to prove its effectiveness in a clinical setting, and its comparative clinical effectiveness with the current standard of care. How to assess DAs nowadays is a matter of research. There are more than 40 assessment frameworks for DAs in an advanced stage of development [65], 21 for eHealth [66], and some author suggestions for types of studies recommended depending on the maturity of the technology [48]. Nevertheless, we did not find any internationally accepted high-quality standard guidance to be used in the design of the WW study which is in an intermediate stage of development (TRL6, TRL ranges from 1–9, a technology with TRL = 9 means that it is in the market). Therefore, we used a study design and an analysis method that better suits the characteristics of WW (i.e., DA, in developmental stage, need for a proof of comparative effectiveness), while approaching the methodological standards required when comparing two technologies for the same purpose. The cross-sectional single-arm clinical trial conducted followed the agile software movement, where incremental, iterative measures are taken, allowing for both empirical feedback and tool improvement [49,67]. This method is one of several that overcomes the challenges that standard methodologies use to test non-digital technologies (e.g., randomized controlled trials with two cohort groups) when they are used to evaluate technologies characterized by a fast evolution and update [20]. This method allowed us to cope with the unprecedented circumstances of COVID-19 during the trial. WW was able to immediately adjust the decision trees and algorithms at the same pace that new COVID symptoms were described. This was important since symptoms associated with COVID-19 were non-specific (i.e., they can be associated with different pathologies), so a proper identification and inclusion in the decision trees, and final classification by level of severity to suggest the right course of action to the user, was needed.

Choosing the right comparator is also a key quality requirement when performing a comparative assessment of technologies [29]. In our study, the diagnosis written in the ED clinical record at discharge was used as the "gold standard" comparator. When reviewing

the clinical records, some non-appropriate reporting was found (i.e., in 45 clinical records, only the symptoms were reported without any diagnosis). Since the end of the 20th century, health information systems have evolved substantially; however, in some settings, a proper coding of diagnosis by doctors still lacks accuracy. If the “gold standard” to compare with the new tool (i.e., WW) has weaknesses, the comparative results may be erroneous, thus penalizing the new tool. This is especially relevant nowadays since real world data (RWD) are more frequently claimed to be used for evaluating technologies. Health systems produce abundant RWD, but its use in research is challenged by both the quality of these data as well as their lack of parameterization. Common data models and standards are still needed to use RWD in a robust and transparent way [68].

To ensure the highest possible quality in the methods used, we explored which statistic was better in assessing WW performance. Traditionally, assessment of concordance of two measurement systems is performed using the Kappa statistic. However, this statistic should be used when the data comply with several assumptions (i.e., normal distribution, a high and similar number of observations between observed agreements and disagreements, marginal values are uniformly distributed) [69]. The assumption that the marginal values are uniformly distributed is not the case in our study, leading to a high weighting of the discordant values when calculating the concordance [47]. This situation underestimates the Kappa results for WW. Several authors investigated how to assess concordance under Kappa limitations [69–72]. These studies compared different estimators of concordance including Cohen’s Kappa, B Statistic, prevalence index (PI), Gweert’s AC1 index (AC1), Matthew, prevalence-adjusted and bias-adjusted Kappa (PABAK), and Delta; it was found that the B statistic gives the closer interpretation to the real concordance (i.e., observed concordance) [69]. The results from our analysis showed that B statistics gets closer to the observed concordance than Kappa. Therefore, the B statistic is a superior statistic to show the real precision of WW as a triage and diagnosis tool.

Several limitations may exist in the assessment of WW comparative effectiveness. We did not use a validated instrument to test WW usability and UX tests. Therefore, the results obtained could differ from another sample of respondents should a validated questionnaire be used. Nevertheless, during the clinical evaluation, minor suggestions were given by users about the WW design and performance. Another limitation during the usability test was the digital literacy measurement. We did not employ a validated questionnaire, opting to use a direct question with several potential answers, which included the range of digital tools most frequently used by society. However, although we would have used a validated questionnaire, all measurements of digitality are subject to change because of technology trends. Moreover, considering the Agency for Health Research and Quality (AHRQ) check list mentioned in Semigran (2015) [73], WW fulfills almost all the requirements to be a health technology suitable for populations with all grades of literacy. Finally, by conducting this first evaluation of WW among patients attending a hospital ED, the results regarding precision and accuracy may only be applicable to those users who have moderate–severe symptoms and not for those with moderate–mild symptoms who must be directed to primary care. Looking at the precision by level of triage, there is a good concordance for the level attending who have moderate–mild symptoms (level IV = B statistic of 0.94), which may indicate that WW triage precision for this type of population is good. However, these interim results should be confirmed in a future study including patients attending emergency primary care.

The promise that WW presents in overcoming both current health system challenges and the inherent limitations of traditional evaluation methods applied to DA technologies such as WW have been previously described and discussed. However, other challenges exist in the race of DAs to reach the market, for example, in legal terms. For any innovative technology, some level of protection to guarantee a high probability of appropriate market access is advisable. Ensuring proper protection for a DA is a key to accessing potential future investments to further develop, operate, and ensure a proper use at DA launching. One of the most reliable forms of intellectual protection is via a patent. However, in the EU,

patentability is not available for most DAs. The lack of a strong protection strategy for these types of technologies renders them at risk to be copied by financially stronger developers or companies. This is especially relevant for start-up companies, one of the main producers of these type of technologies and who typically have tight budgets. Regulation represents another central challenge. The new EU medical device regulation (MDR) considers DAs to be a medical device (MD) [27]. However, WW is placed in a gray area when considering the definition of MDs by the new MDR. WW does not provide the user with a diagnosis, which is solely stored in the system. Therefore, although it is a tool linked to health care, it may not be considered strictly as a MD by some NB. In fact, one NB in the EU considered WW not to be a MD, instead considering it to be in a gray zone. It could be the case that other NBs consider WW as a MD. Not having a regulatory granting may be detrimental for the product since potential users are asking for some form of quality certification before using this type of product [74].

Besides the CE mark, compliance with other regulatory requirements (i.e., ISO13485) is needed in the EU for software. Reaching the CE mark and obtaining and maintaining an ISO is extremely difficult, requiring intensive human and economic resources. A question to ponder is whether the “one-size-fits-all model” that is currently asked for any type of MD, under the new MDR, should still be kept for emerging DAs, or if a new model specifically designed for a DA that stratifies regulatory requirements by type of digital technology (including DAs) and technology readiness level (TRL) of the product should be created.

Finally, another challenge for innovative DAs are the EU evidentiary requirements by public payers. In Europe, having a CE mark for a MD is necessary for market entrance; however, it does not guarantee adoption by healthcare systems. Proving that the MD performs at least equally or better than the standard of care is demanded by payers. Currently, there is no clear, standardized guidance on how to clinically assess the wide range of DAs. As already mentioned, a total of 45 frameworks to assess mobile applications for medicine have been identified by a systematic review [65]. Moreover, different challenges have been described in applying the traditional clinical research methods to DAs [20]. In the search to look for a more appropriate process to assess DAs, Germany has recently implemented the DiGA process for DAs with a CE mark class I and IIa. DiGA was born as a fast-track to market for those DAs with demonstrated benefits or with great potential. When a DA enters DiGA, the reimbursement is guaranteed to prove that the benefits claimed are obtained after one year of implementation. At the end of this period, a new reimbursement level is negotiated or withdrawn in the event that the DA fails to prove beneficial. DiGA does not have explicit public methodological guidance to assess the DA; methods and outcomes to be assessed are set up case by case through discussions between the DA producer and BfARM, the regulatory agency for MD in Germany [75]. While DiGA is a good first step to help advance a DA to market, the short experience shows that several weaknesses in the model exist [76]. Other initiatives also exist in Europe to guide the assessment of DA to obtain reimbursement. One is the so called “Belgium pyramid” that increases the complexity of scientific requirements that prove DA benefits linked to the level of its classification (class I, II, III) [77]. Similarly, the National Institute for Health and Clinical Excellence (NICE) has recently issued a new version of its Evidence Standard Framework (ESF) for digital health technologies, which classifies technologies by their potential risk to service users and systems according to their intended purposes. Depending on the classified tier, the DA will have to comply with different evidentiary requirements [78]. Here, a challenge may appear for WW due to a contradictory position between a notified body (which considers WW not to be a MD) and NICE which, through its ESF, would consider WW as a MD and classify it in tier 3, asking for the most comprehensive and strict fulfillment of evidentiary requirements. As such, the evidence generation approaches that are appropriate for DAs, to fulfill most health authorities’ requirements, is a theme of current debate in European member states. The methods used in our study for assessing the comparative effectiveness of WW, following suggestions

from other authors [49,53], may be used as an inspirational insight for researchers and health technology assessment scientists in developing more appropriate methods for the evaluation of digital technologies at different TRLs.

5. Conclusions

Digital health is among the key phenomena driving the next cycle of transformation of health systems. Digital technologies—DA, wearables, and software algorithms—have the potential to offer easier and more agile access to support a technology-enabled health system in which care interactions with citizens, for some medical encounters, are moved away from traditional settings by encouraging citizens to manage their own health and mild health problems. In turn, clinicians will optimize their time using artificial intelligence (AI) solutions as a support to their daily work. Accordingly, digital health will enable “using humans for the hard stuff and leaving the basics to machines” [79].

Nevertheless, the proper development and clinical testing of the benefits of emerging DAs is a challenge in traditional healthcare systems. Weaknesses in the available RWD of comparators, lack of strong protection rules in some jurisdictions, requirements to use traditional study designs for technologies that are continuously evolving, and/or lack of accepted international standards for granting the access and coverage for DAs into health systems are obstacles to developing innovative DAs, especially for start-ups who are the main developers of these types of technologies.

WW is a promising and disruptive DA, with a high potential scalability and social impact that will allow citizens worldwide to access advice and answers on what to do when faced with a basic health concern. To prove its value from inception, it has been required to overcome challenges which characterize most current health systems (i.e., information system design and content, regulatory and assessment evidentiary requirements) with novel approaches for evidence generation that allows for its agile, and patient-centered, development. Though still not marketed, and in need of advancement in terms of WW maturity level, the experience developing WW should be a matter of critical, reflective thinking for DA developers, digital health scientists, regulators, health technology assessors, and payers, in the quest for designing an appropriate pathway to a market tailored to DAs.

Author Contributions: Conceptualization: L.S.-C.; Methodology: C.F.-B., I.A., B.G.-L., L.S.-C. and A.G.; Software: N.R.; Validation: C.F.-B., I.A., L.S.-C. and A.G.; Formal analysis: C.F.-B., I.A. and B.G.-L.; Investigation: L.S.-C., C.F.-B., B.V., N.R., A.G., M.S. (Mireia Sans), B.G.-L., M.S. (Miquel Sanchez) and N.G.; Data curation: C.F.-B., B.V. and N.R.; Writing—original draft preparation: L.S.-C.; Writing—review and editing: C.F.-B., A.G., M.S. (Mireia Sans), B.G.-L., M.S. (Miquel Sanchez) and N.G.; Project administration: L.S.-C.; Supervision: L.S.-C. and A.G.; Funding Acquisition: L.S.-C. and A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been co-funded by the Clinic Barcelona University Hospital (Spain), Mangrana Ventures S.L., and EIT Health Spain (European Institute of Technology Health) under its Headstart Program (EIT Health Activity HS PoC 2019-HS-0038).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Clinic Barcelona University Hospital (protocol code Reg. CB/2020/0087 and 9 July 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to patient data protection.

Acknowledgments: The authors would like to thank the contribution of Montserrat Ramos for helping with the organization of the clinical validation process at the emergency department of the Clínic Barcelona University Hospital. The authors would also like to thank Artur Conesa and Laura Faedo from the Medical Informatics Documentation Unit at the Clínic Barcelona University Hospital for their guidance on the data retrieval process and assistance in understanding it. Thank you also to Ana Patitucci for developing the images of WtsWrng. We also wish to thank the initial investors and the advisory board of Mangrana Ventures SL for their economic support in developing WtsWrng.

Conflicts of Interest: Author Laura Sampietro-Colom, Alicia Granados, and Nicolau Rabasseda are co-founders of the start-up Mangrana Ventures S.L., developer of the DA WtsWrng presented in this paper. Alicia Granados is also Chair of Scientific Committee of Mangrana S.L.

References

1. Jones, R. Access to primary care: Creative solutions are needed. *Br. J. Gen. Pract.* **2015**, *65*, 620–621. [CrossRef] [PubMed]
2. World Health Organization. World Bank and WHO: Half the World Lacks Access to Essential Health Services, 100 Million Still Pushed into Extreme Poverty Because of Health Expenses. Available online: <https://www.who.int/news/item/13-12-2017-world-bank-and-who-half-the-world-lacks-access-to-essential-health-services-100-million-still-pushed-into-extreme-poverty-because-of-health-expenses> (accessed on 17 August 2022).
3. World Health Organization. Tracking Universal Health Coverage: 2017 Global Monitoring Report. 2017. Available online: <https://apps.who.int/iris/bitstream/handle/10665/259817/9789241513555-eng.pdf> (accessed on 17 August 2022).
4. Bresnick, J. How Many Emergency Department Visits Are Really Avoidable? 2017. Available online: <https://healthitanalytics.com/news/how-many-emergency-department-visits-are-really-avoidable> (accessed on 17 August 2022).
5. Weinick, R.M.; Burns, R.M.; Mehrotra, A. Many Emergency Department Visits Could be Managed at Urgent Care Centers and Retail Clinics. *Health Aff.* **2010**, *29*, 1630–1636. [CrossRef] [PubMed]
6. Renfrow, J. UnitedHealth: Healthcare Could Save \$32B Annually by Diverting Avoidable ED Visits | Fierce Healthcare. 2019. Available online: <https://www.fiercehealthcare.com/payer/two-thirds-ed-visits-avoidable> (accessed on 31 August 2022).
7. Zane, R. The Hardest Time in the History of Emergency Medicine. *NEJM Catal. Innov. Care Deliv.* **2022**, *3*. Available online: <https://catalyst.nejm.org/doi/full/10.1056/CAT.22.0009> (accessed on 29 August 2022). [CrossRef]
8. Schwartz, E. The Global Health Care Worker Shortage: 10 Numbers to Note. Available online: <https://www.projecthope.org/the-global-health-worker-shortage-10-numbers-to-note/04/2022/> (accessed on 23 August 2022).
9. Sonis, J.D.; Aaronson, E.L.; Lee, R.Y.; Philpotts, L.L.; White, B.A. Emergency Department Patient Experience: A Systematic Review of the Literature. *J. Patient Exp.* **2018**, *5*, 101–106. [CrossRef] [PubMed]
10. Soler, W.; Gómez Muñoz, M.; Bragulat, E.; Álvarez, A. El triaje: Herramienta fundamental en urgencias y emergencias. *An. Sist. Sanit. Navar.* **2010**, *33*, 55–68. [CrossRef]
11. Bullard, M.J.; Musgrave, E.; Warren, D.; Unger, B.; Skeldon, T.; Grierson, R.; van der Linde, E.; Swain, J. Revisions to the Canadian Emergency Department Triage and Acuity Scale (CTAS) Guidelines 2016. *CJEM* **2017**, *19*, S18–S27. [CrossRef]
12. Robertson-Steel, I. Evolution of triage systems. *Emerg. Med. J.* **2006**, *23*, 154–155. [CrossRef]
13. Joint Commission International. Communicating Clearly and Effectively to Patients How to Overcome Common Communication Challenges in Health Care. 2018. Available online: [https://store.jointcommissioninternational.org/assets/3/7/jci-wp-communicating-clearly-final_\(1\).pdf](https://store.jointcommissioninternational.org/assets/3/7/jci-wp-communicating-clearly-final_(1).pdf) (accessed on 30 August 2022).
14. Allar, B.; Ortega, G. Our Health System Is Failing Patients with Limited English—Scientific American. *Scientific American*. 2021. Available online: <https://www.scientificamerican.com/article/our-health-system-is-failing-patients-with-limited-english/> (accessed on 29 August 2022).
15. Allen, S. 2020 Global Health Care Outlook. Laying a Foundation for the Future. Deloitte Insights. 2020. Available online: <https://www2.deloitte.com/content/dam/insights/us/articles/GLOB22843-Global-HC-Outlook/DI-Global-HC-Outlook-Report.pdf> (accessed on 23 January 2020).
16. Kirk, A. One in Four Self-Diagnose on the Internet Instead of Visiting the Doctor. 2015. Available online: <https://www.telegraph.co.uk/news/health/news/11760658/One-in-four-self-diagnose-on-the-internet-instead-of-visiting-the-doctor.html> (accessed on 29 August 2022).
17. Brucher, L.; Pontet, K.; Duprel, C.; Mosel, K.; Moujahid, S.; Georges, A. A Journey towards Smart Health: The Impact of Digitalization on Patient Experience. 2018. Available online: https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/life-sciences-health-care/lu_journey-smart-health-digitalisation.pdf (accessed on 13 January 2020).
18. Meskó, B.; Dhunoo, P. Digital Health and the Fight against the COVID-19 Pandemic. 2020. Available online: <https://www.matrc.org/wp-content/uploads/2020/04/Digital-Health-and-COVID19.pdf> (accessed on 29 August 2022).
19. Mehrotra, A.; Ray, K.; Brockmeyer, D.M.; Barnett, M.L.; Bender, J.A. Rapidly Converting to “Virtual Practices”: Outpatient Care in the Era of COVID-19. 2020. Available online: <https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0091> (accessed on 29 August 2022).
20. Guo, C.; Ashrafian, H.; Ghafur, S.; Fontana, G.; Gardner, C.; Prime, M. Challenges for the evaluation of digital health solutions—A call for innovative evidence generation approaches. *NPJ Digit. Med.* **2020**, *3*, 110. [CrossRef]
21. Greaves, F.; Joshi, I.; Campbell, M.; Roberts, S.; Patel, N.; Powell, J. What is an Appropriate Level of Evidence for a Digital Health Intervention? *Lancet* **2018**, *392*, 2665–2667. [CrossRef]
22. Chambers, D.; Cantrell, A.J.; Johnson, M.; Preston, L.; Baxter, S.K.; Booth, A.; Turner, J. Digital and online Symptom Checkers and Health Assessment/Triage Services for Urgent Health Problems: Systematic Review. *BMJ Open* **2019**, *9*, e027743. [CrossRef]
23. Atluri, V.; Rao, S.; Rajah, T.; Schneider, J.; Thibaut, M.; Varanasi, S.; Velamoor, S. *Unlocking Digital Health: Opportunities for the Mobile Value Chain*; McKinsey & Company: Atlanta, GA, USA, 2015; p. 8. Available online: <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/unlocking-digital-health-opportunities-for-the-mobile-value-chain> (accessed on 9 December 2019).

24. Yu, S.W.Y.; Ma, A.; Tsang, V.H.M.; Chung, L.S.W.; Leung, S.C.; Leung, L.P. Triage Accuracy of Online Symptom Checkers for Accident and Emergency Department Patients. *Hong Kong J. Emerg. Med.* **2019**, *27*, 217–222. [CrossRef]
25. Verzantvoort, N.C.M.; Teunis, T.; Verheij, T.J.M.; van der Velden, A.W. Self-triage for acute primary care via a smartphone application: Practical, safe and efficient? *PLoS ONE* **2018**, *13*, e0199284. [CrossRef]
26. Christensen, C.M.; Bohmer, R.M.J.; Kenagy, J. Will Disruptive Innovations Cure Health Care? *Harv. Bus. Rev.* **2000**, *78*, 102–112.
27. European Union Medical Device Regulation. The European Union Medical Device Regulation—Regulation (EU) 2017/745 (EU MDR). 2022. Available online: <https://eumdr.com/> (accessed on 19 August 2022).
28. O'Rourke, B.; Oortwijn, W.; Schuller, T. The New Definition of Health Technology Assessment: A Milestone in International Collaboration. *Int. J. Technol. Assess. Health Care* **2020**, *36*, 187–190. [CrossRef]
29. EUnetHTA. HTA Core Model. 2016. Available online: <https://www.eunethta.eu/hta-core-model/> (accessed on 7 September 2022).
30. Millenson, M.L.; Baldwin, J.L.; Zipperer, L.; Singh, H. Beyond Dr. Google: The Evidence on Consumer-Facing Digital Tools for Diagnosis. *Diagnosis* **2018**, *5*, 95–105. [CrossRef]
31. Sole, M.L.; Stuart, P.L.; Deichen, M. Web-Based Triage in a College Health Setting. *J. Am. Coll. Health* **2010**, *54*, 289–294. [CrossRef]
32. Salisbury, C.; Procter, S.; Stewart, K.; Bowen, L.; Purdy, S.; Ridd, M.; Valderas, J.; Blakeman, T.; Reeves, D. The Content of General Practice Consultations: Cross-Sectional Study based on Video Recordings. *Br. J. Gen. Pract.* **2013**, *63*, e751–e759. [CrossRef] [PubMed]
33. Duckitt, R.; Palsson, R.; Bosanska, L.; Dagna, L.; Mine, D.T.; Vardi, M. Common diagnoses in Internal Medicine in Europe 2009: A pan-European, Multi-Centre Survey. *Eur. J. Intern. Med.* **2010**, *21*, 449–452.
34. Walls, R.M.; Hockberger, R.S.; Gausche-Hill, M. *Rosen's Emergency Medicine: Concepts and Clinical Practice*, 10th ed.; Elsevier: Amsterdam, The Netherlands, 2022; 148p.
35. Stern, S.D.C.; Cifu, A.S.; Altkorn, D. *Symptom to Diagnosis: An Evidence-Based Guide*, 4th ed.; McGraw Hill Medical: New York, NY, USA, 2019. Available online: <https://accessmedicine.mhmedical.com/book.aspx?bookID=2715> (accessed on 18 August 2022).
36. Guillén, B.C. *Manual de Urgencias*, 3rd ed.; Coordinadores y SANED: Madrid, Spain, 2018; pp. 1–1653.
37. Stanford Anesthesia Cognitive Aid Program. Emergency Manual. 2022. Available online: http://web.stanford.edu/dept/anesthesia/em/SEM_printing.pdf (accessed on 19 August 2022).
38. Family Doctor. Symptom Checker. Available online: <https://familydoctor.org/your-health-resources/health-tools/symptom-checker/> (accessed on 18 August 2022).
39. Simon, C.; O'Reilly, K.; Proctor, R.; Buckmaster, J. *Emergencies in Primary Care*; Oxford University Press: Oxford, UK, 2007. Available online: <https://academic.oup.com/book/31748> (accessed on 18 August 2022).
40. Campo, T.M.; Lafferty, K.A. *Essential Procedures for Emergency, Urgent, and Primary Care Settings*, 3rd ed.; Springer Publishing Company: New York, NY, USA, 2021.
41. Rare Diseases Registry Program. Single Group Study. Available online: <https://registries.ncats.nih.gov/glossary/single-group-study/> (accessed on 6 October 2022).
42. Clinicaltrials.gov. Transversal, Controlled and Not Randomized Study, Laying on the Desire of Children and Eating Disorders in a Population of Infertile Women. Available online: <https://clinicaltrials.gov/ct2/show/NCT02029365> (accessed on 6 October 2022).
43. Health Research Authority. Fast Track Review Guidance for COVID-19 Studies. 2022. Available online: <https://www.hra.nhs.uk/covid-19-research/fast-track-review-guidance-covid-19-studies/> (accessed on 30 August 2022).
44. Jiménez, J.G. Modelo Andorrano De Triaje (Model Andorrà De Triatge: Mat). 2003. Available online: <https://docplayer.es/20789909-Modelo-andorrano-de-triaje-model-andorra-de-triatge-mat.html> (accessed on 18 August 2022).
45. Emergency Nurses Association. *Emergency Severity Index: A Triage Tool for Emergency Department Care*; Emergency Nurses Association: Schaumburg, IL, USA, 2020.
46. Carrasco, L.J.; Jover, L. Métodos Estadísticos Para Evaluar la Concordancia. *Med. Clin.* **2004**, *122*, 28–34. [CrossRef]
47. Shankar, V.; Bangdiwala, S.I. Behavior of Agreement Measures in the Presence of Zero Cells and Biased Marginal Distributions. *J. Appl. Stat.* **2008**, *35*, 445–464. [CrossRef]
48. Jansen-Kosterink, S.; Broekhuis, M.; van Velsen, L. Time to Act Mature—Gearing eHealth Evaluations towards Technology Readiness Levels. *Digit. Health* **2022**, *8*, 205520762211133. [CrossRef] [PubMed]
49. Martinez, B.; Hall-Clifford, R.; Coyote, E.; Stroux, L.; Valderrama, C.E.; Aaron, C.; Francis, A.; Hendren, C.; Rohloff, P.; Clifford, G.D. Agile Development of a Smartphone App for Perinatal Monitoring in a Resource-Constrained Setting. *J. Health Inform. Dev. Ctries* **2017**, *11*. Available online: <http://www.jhidc.org/index.php/jhidc/article/view/158/212> (accessed on 30 August 2022).
50. StataCorp. *Stata Statistical Software: Release 13*; StataCorp LP: College Station, TX, USA, 2013.
51. IPA4SME Consortium. IPA4SME. 2019. Available online: <https://www.ipa4sme.eu/> (accessed on 30 August 2022).
52. IBM. What is Supervised Learning? IBM Cloud Education. 2020. Available online: <https://www.ibm.com/cloud/learn/supervised-learning> (accessed on 31 August 2022).
53. Christensen, C.; Waldeck, A.; Fogg, R. How Disruptive Innovation Can Finally Revolutionize Healthcare: A Plan for Incumbents and Startups to Build a Future of Better Health and Lower Costs. Industry Horizons. 2017. Available online: <https://www.christenseninstitute.org/wp-content/uploads/2017/05/How-Disruption-Can-Finally-Revolutionize-Healthcare-final.pdf> (accessed on 29 August 2022).

54. Statista. Smartphone Subscriptions Worldwide 2027. 2022. Available online: <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/> (accessed on 7 September 2022).
55. Miller, J. Checking Up on Symptom Checkers. Harvard Medical School. 2015. Available online: <https://hms.harvard.edu/news/checking-symptom-checkers> (accessed on 29 August 2022).
56. Ipsos, M.; York Health Economics Consortium; Salisbury, C. Evaluation of Babylon GP at Hand Final Evaluation Report. 2019. Available online: <http://www.ipsos-mori.com/terms> (accessed on 29 August 2022).
57. Wallace, W.; Chan, C.; Chidambaram, S.; Hanna, L.; Iqbal, F.M.; Acharya, A.; Normahani, P.; Ashrafian, H.; Markar, S.R.; Sounderajah, V.; et al. The Diagnostic and Triage Accuracy of Digital and Online Symptom Checker Tools: A Systematic Review. *NPJ Digit. Med.* **2022**, *5*, 1–9. [CrossRef]
58. Ceney, A.; Tolond, S.; Glowinski, A.; Marks, B.; Swift, S.; Palser, T. Accuracy of Online Symptom Checkers and the Potential Impact on Service Utilisation. *PLoS ONE* **2021**, *16*, e0254088. [CrossRef]
59. Eisenhauer, T. Top 10 Business Communication Trends for 2022—Axero Solutions. Available online: <https://axerosolutions.com/blog/top-10-business-communication-trends-improve-communication-in-the-workplace> (accessed on 24 August 2022).
60. Statista. Video Games—Worldwide | Statista Market Forecast. 2022. Available online: <https://www.statista.com/outlook/dmo/digital-media/video-games/worldwide> (accessed on 30 August 2022).
61. Cheung, M.M.Y.; Saini, B.; Smith, L. Using drawings to explore patients’ perceptions of their illness: A scoping review. *J. Multidiscip. Healthc.* **2016**, *9*, 631. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5125999/> (accessed on 6 October 2022). [CrossRef]
62. Delamerced, A. The Importance of Medical Illustration in Patient Communication. In-Training. 2018. Available online: <https://in-training.org/importance-medical-illustration-patient-communication-15851> (accessed on 6 October 2022).
63. Deloitte Centre for Health Solutions. Connected Health—How Digital Technology Is Transforming Health and Social Care. Deloitte. 2015, p. 40. Available online: <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/deloitte-uk-connected-health.pdf> (accessed on 5 December 2019).
64. Moshi, M.R.; Parsons, J.; Tooher, R.; Merlin, T. Evaluation of Mobile Health Applications: Is Regulatory Policy Up to the Challenge? *Int. J. Technol. Assess Health Care* **2019**, *35*, 351–360. [CrossRef]
65. Moshi, M.R.; Tooher, R.; Merlin, T. Suitability of current evaluation frameworks for use in the health technology assessment of mobile medical applications: A systematic review. *Int. J. Technol. Assess Health Care* **2018**, *34*, 464–475. [CrossRef]
66. Vis, C.; Bührmann, L.; Riper, H.; Ossebaard, H.C. Health technology assessment frameworks for eHealth: A systematic review. *Int. J. Technol. Assess Health Care* **2020**, *36*, 204–216. [CrossRef]
67. Principles behind the Agile Manifesto. Available online: <http://agilemanifesto.org/principles.html> (accessed on 30 August 2022).
68. Oortwijn, W.; Sampietro-Colom, L.; Trowman, R. How to Deal with the Inevitable: Generating Real-World Data and Using Real-World Evidence for HTA Purposes—From Theory to Action. *Int. J. Technol. Assess Health Care* **2019**, *35*, 346–350. [CrossRef]
69. Gwet, K. Kappa Statistic Is Not Satisfactory for Assessing the Extent of Agreement between Raters. *Statistical Methods for Inter-Rater Reliability Assessment*. 2002, pp. 1–16. Available online: [https://www.scirp.org/\(S\(lz5mqp453edsnp55rrgjt55.\)\)reference/referencespapers.aspx?referenceid=2208179](https://www.scirp.org/(S(lz5mqp453edsnp55rrgjt55.))/reference/referencespapers.aspx?referenceid=2208179) (accessed on 17 August 2022).
70. Shankar, V.; Bangdiwala, S.I. Observer agreement paradoxes in 2×2 tables: Comparison of agreement measures. *BMC Med. Res. Methodol.* **2014**, *14*, 100. [CrossRef]
71. Delgado, R.; Tibau, X.A. Why Cohen’s Kappa should be avoided as performance measure in classification. *PLoS ONE* **2019**, *14*, e0222916. [CrossRef]
72. Dettori, J.R.; Norvell, D.C. Kappa and Beyond: Is There Agreement? *Glob. Spine J.* **2020**, *10*, 499–501. [CrossRef]
73. Semigran, H.L.; Linder, J.A.; Gidengil, C.; Mehrotra, A. Evaluation of symptom checkers for self diagnosis and triage: Audit study. *BMJ* **2015**, *351*, h3480. [CrossRef]
74. IQVIA. Digital Health Trends. 2021. Available online: <https://www.iqvia.com/insights/the-iqvia-institute/reports/digital-health-trends-2021> (accessed on 30 August 2022).
75. EIT Health Germany. DiGAs—A Model for Europe? Possible Options for Achieving a European System. 2021. Available online: <https://eithealth.eu/wp-content/uploads/2022/03/The-DiGAs-framework-a-model-for-Europe.pdf> (accessed on 29 August 2022).
76. Jungman, S.; Kraft, D. European Healthtech Needs Stronger Medicine Than Germany’s Weak Digital Law. 2022. Available online: <https://sifted.eu/articles/europe-germany-digital-health/> (accessed on 30 August 2022).
77. MedTech Europe. Recognising the Value of Digital Health Apps: An Assessment of Five European Healthcare Systems. 2021. Available online: www.medtecheurope.org (accessed on 30 August 2022).
78. National Institute for Health and Care Excellence. *Evidence Standards Framework for Digital Health Technologies Contents*; National Institute for Health and Care Excellence: London, UK, 2019.
79. Zane, R.D.; Wiler, J.L. Embracing Technology to Save Primary Care. *N. Engl. J. Med. Catal.* **2018**. Available online: <https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0101> (accessed on 30 August 2022).