Table S1. Overview of the manuscripts included in the systematic review.

Authors	App	App Aim	Individuals Involved In The Usability Assessment	Use Case	Domain Assessed	Method Of Usability Assessment Used	Usability Outcome	Reviewers Comments
Dantas_2016 [25]	Dor de Cabeça; Diário da Dor; Diário da Cefaleia	Assessment	End users N = 22 Characteristics 54.5% (n=12) female and 45.5% (n=10) males; mean (±SD) age of 72.3 (±8.2) years old	Not specified	Efficiency, Effectiveness Satisfaction	International Classification of Functioning - Usability Scale I (ICF-US I); Performance evaluation: i) Number of attempts to complete tasks ii) Completion rates iii) Time taken to complete tasks	ICF results: Dor de Cabeça (Mean±sd ICF_US I=25.4 ±0.4), Diário da Dor (Mean±sd ICF_US I=24.7±3.9) Diário da Cefaleia (Mean±sd ICF_US I= 13.6±3.7) Time to complete tasks: Diário da Cefaleia (mean±sd=7.1±3.3min) Dor de Cabeça (mean ±sd=6.2 ±2.9 min) Diário da Dor (mean±sd=5.1 ±2.7 min)	
Reynoldson_2014 [26]	1. Manage My Pain; 2. Pain Scale	Assessment	End users N = 41 (aged 19–59 years, with experience of chronic or recurrent pain)	Record 2 pain episodes (average level of pain they regularly experience and worst pain they have experienced)	Efficiency Effectiveness Satisfaction	1. Time taken to complete a data entry task on the app; 2. SUS 3. Design questionnaire; 4. App use questionnaire	Usability task timings Manage My Pain - ranged from 1′22″ to 6′09″ (mean = 3′01″; SD= 1′14″) for the first entry and from 0′44″ to 4′05″ (mean = 2′00″; SD= 0′49″) for the second entry. Pain Scale app - ranged from 0′02″ to 0′59″ (mean = 0′19″;SD= 0′12″) for the first entry and from 0′02″ to 0′37″ (mean = 0′14″; SD= 0′08″) for the second entry. SUS ManageMyPain scores ranged from 57.5 to 100. SUS PainScale scores ranged from 60 to 100.	
Spyridonis_2014 [28]	PainDroid	Assessment	End users N = 7 wheelchair users (3 fem; 4 males) with some	1. Start the app	Satisfaction	Questionnaire adapted from SUS	Positive trends in the questions 2 and 3, regarding ease of use.	Total score of the SUS adapted

			arm mobility problems, ranging in age from 21 to 65 years (mean age 41.1) Experts N = 2 clinicians (a General Practitioner-GP- and a Rheumatologist-RH)	2. Rotate left/right, rotate up/down 3. Re-center 4. Zoom in/out and drag 5. Reset 6. Use above to select pain type and pain location on model 7. Save and exit			There was strong disagreement with statements targeting the amount of learning required to have in order to use PainDroid.	questionnaire is not presented
Jibb_2017 [29]	Pain Squad+	Intervention	End users N = 16 Cycle 1 (n = 4) Cycle 2 (n = 6), Cycle 3 (n = 6) Characteristics Adolescents with cancer, ranged in age from 12 to 18 years (mean = 14.8; SD = 2.1), 9 female, 7 male.	Adolescents were asked to complete the app (pain assessment, pain management, and gamification mechanics)	Satisfaction Efficiency Effectiveness	Think aloud approach Audio-recording and field notes Semi-structured interview Recording of time to complete the pain assessment	All participants but one were able to complete the tasks in a single session. Participants across all testing cycles stated that Pain Squad+ was easy to use. The time to complete the pain assessment component of the app across all iterative cycles was 4.3 ± 3.5 minutes (M ± SD; range = 1.9-14.3 minutes). All participants across the 3 cycles were satisfied with the Pain Squad+ app.	
Jibb_2018 [30]	Pain Squad+	Intervention	End users N = 20 Characteristics Adolescents with cancer (12-18 years)	Adolescents were asked to complete the app	Satisfaction Effectiveness	Count number of registers on the app Semi-structured telephone interview audio-recorded, with field notes	95% (n=19) of the participants completed the entire pilot study. Every adolescent endorsed the ease of use of Pain Squad+. All participants endorsed pain management advice, design and gamification mechanics.	
De Knegt_2016 [31]	STOP-ID!	Assessment	End users N = 40 Characteristics	1) Patients were asked if they were currently	Efficiency, Effectiveness	Qualitative observations	All participants finished the STOP-ID! The average performance time was 14.6 min (SD¼3.9)	

			Adults with Down syndrome, average age of 43.3 years (SD¼11.7, range: 20–66 years)	experiencing pain; 2) locate the pain; 3) classify pain intensity; 4) assess pain affect selecting Pictograms; 5) assess pain Quality selecting Pictograms		Performance evaluation following a form (e.g., presence of distraction or impulsivity); Time to complete the test; Number of times that the instructions needed to be repeated	48% of the participants seemed at times distracted or bored and, 35% appeared to answer at least a few questions impulsively 7 participants were able to use the tool independently, 4 needed assistance with the device, 16 needed assistance in the questions about the comprehension tests and/or self-report of pain, and 13 needed assistance with both the features. Verbal and/or non-verbal requests of assistance were made by 70% of the participants.
Kaltenhauser_2018 [32]	Quiri	Assessment	End users N = 3 adults Experts N = 3	Three tasks (not specified)	Effectiveness	Successful completion of tasks Discussions with physicians	All participants completed the three tasks. The ease of use and short interaction time to complete a measurement were highlighted as particularly valuable by the physicians.
Jaatun, E_2013 [33]	Pain Body Map (PBM)	Assessment	End users N = 10 Characteristics Patients with advanced cancer (both in the first and second versions of the iPad PBM)	1) Mark the area of pain in the body map, 2) Select the intensity of the pain by selecting the image of the corresponding numbered colouring pencil/radio button.	Effectiveness	Observation of task execution – completion rates	First version 3 of the 10 patients had no problem using the iPad and complete tasks; 5 were able to mark intensity but needed orientation; 2 patients were not able to give any input on the iPad (one was too frail and the other was unable to follow instructions). Second Version

							8 participants completed the tasks without support; 2 patients were not able to fill in anything at all due to drowsiness /sleepiness and having problems following instructions.	
Diana_2012 [34]	No reference to app name	Assessment	End users N = 5 (woman with fibromyalgia)	Not specified	Satisfaction Effectiveness Efficiency*	Mock-ups of the app Questionnaire Recording of the sessions Success/failure rates Time to complete tasks	All the participants were able to give ratings in the three scales. 100% of the users found the device was easy to use 75% of the users thought they would be able to use the device. 75% of the users had difficulties to understand how to advance (did not understand the metaphor of the button, represented by an arrow.	The authors only present the mean time of each session, they make no reference to the time to complete the tasks
Caon_2019 [35]	My Pain Coach	Assessment	End users N = 6 Characteristics People without any chronic pain condition	1) Register; 2) create a new "pain entry" with Tangible Interface; 3) create a new "pain entry" with Touch screen; 4) View reports and visualize "Pain map"; 5) Insert a password for doctors.	Satisfaction Efficiency Effectiveness	SUS Recording of time to complete tasks Completion rates of the tasks	The SUS score assessed usability as "excellent" (mean result of all users was 82,5±5,7); users perceived this system as easy-to-use. Time to complete tasks (in seconds) ± SD (mean of the 6 users): task1 – 141±72; task2 – 229±71; task3 - 153±55; task4 - 31±14 task5 - 33±13 Success rate - 100% (all participants completed all tasks)	

Birnie, K_2018 [36]	Achy Penguin	Intervention	End users N = 20 (Cycle 1, n=6, Cycle 2 and 3, n= 7) Characteristics 4 to 7 year-old children (mean age 5.8 years) Experts N = 2 Characteristics Child life specialists at Seattle Children's Hospital reviewed and tested the initial app prototype	The participants progressed through all aspects of the app, including pain assessment (pain location and intensity) and pain management strategies	Satisfaction Efficiency Effectiveness	Qualitative interview; Think aloud approach (likes, dislikes, and difficulties using the app); Field notes on children's responses (taken by research assistants); Recording of the length of time children engaged with app features.	Time engagement with app features: Pain intensity (M, SD, range in seconds): Cycle1 - 44.2, ±22.9, 30–90 Cycle2 - 21.7, ±20.1, 10–60 Cycle3 - 24.7, ±18.2, 10–60 Pain location (M, SD, range in seconds): Cycle1 - 90.0, ±53.7, 30–180 Cycle2 - 23.6, ±18.4, 15–60 Cycle3 - 40.0, ±30.3, 10–90 Self-management strategies (M, SD, range in minutes): Cycle1 – 7.7, ±2.7, 3–10 Cycle2 - 9.1, ±3.5, 6–16 Cycle3 - 14.0, ±1.8, 11–16
Sun_2018 [37]	Panda	Assessment	End users N = 17 • 12 parents (4 in round 1, 3 in round 2, 5 in round 3), • 5 adolescents (round 3) Experts N = 13 nurses (6 in round 1, 7 in round 2)	a) Enter child and scheduled medication b) Respond to app alert, performing a pain assessment and complete checks for medication administration, and then edit the status of the medication c) Respond to a notification when a medication has already been given,	Efficiency Effectiveness Satisfaction	1) Task errors rate 2) Task completion rate 3) Time to complete tasks 4) "think aloud" and was audio-recording of the sessions. 5) Computer System Usability Questionnaire (CSUQ) 5) Unstructured interview	All users were able to successfully complete tasks a) and b). Nurses tended to perform better in terms of task completion rate and time, as well as task error rates Users' CSUQ responses had median [IQR] score of 2 [1-4]; 67% of users indicated that they would use Panda for management of postoperative pain. 93% of participants reported the app was easy to use

				d) Schedule medication, edit and delete existing medication.				
Yen_2016 [38]	RhEumAtic Disease activitY (READY)	Assessment	End users N = 33 patients N = 15 physicians	1) Completing questionnaires, 2) Navigate the system, 3) Patient and physician assessment of pain 4) Documenting medications, 5) interpret results	Satisfaction Effectiveness Efficiency	1) Think Aloud protocol 2) Video-recording of interactions with READY 3) Comparison of the time using READY and paper evaluation.	Usability issues were found in touchscreen sensitivity, interface design, interactive features, and instruction and error messages. Despite these issues, 75.76% patients reported they liked READY. About 37% of patients found touch screen sensitivity was not sensitive enough to complete an action. Physicians faced similar issues. Comparing the times, results show that patients spent more time on READY than paper (mean=4.39 mins, SD=2.29 vs. mean=2.26 mins, SD=1.36, p=0.002). In this matter, physicians also revealed concerns about clinical workflow.	
De La Veja_2014 [39]	Painometer	Assessment	End users N = 14 (non-professionals, with history of pain) Experts N = 19 (professionals)	Assessing pain intensity with the scales; see list of scales; see description of the scale and instructions; choose or change scale.	Satisfaction Effectiveness	Semi-structured interview Open-ended questions Recording of errors Recording of sessions and field notes	Most of health care professionals (n=18, 95%) preferred Painometer to traditional versions of the scales. Participants completed the tasks, but 3 (1 non-professional and 2 professionals) had some difficulties with the buttons size.	Results of the task errors are not presented

						Tasks completion rates Think Aloud approach		
De La Veja_2018 [40]	Fibroline	Intervention	End users N = 25 Characteristics 13 to 24 years (mean age = 18.24; SD = 4.02), 8 males (32%) and 17 females (68%) with fibromyalgia	Log in the app, change settings, access certain resources—read a PDF file, listen to an audio, and watch a video, set an alarm for medication, go back to the main screen	Satisfaction	Semi structured interview (Yes/No questions and open-ended questions) Field notes of the sessions and recording of errors Think Aloud approach	All participants (100%) rated Fibroline as "Easy to use." All participants (100%) would use Fibroline if they needed it. Problems reported using the app - finding button to access chart, setting alarm, choose to see some content in text or video format.	The results of the recorded mistakes are not presented.
Suso-Ribera_2018 [41]	Pain Monitor	Assessment	End Users N = 38 (patients with chronic pain; mean age 42.74, SD = 9.92) Experts N = 7 (two psychologists and five physicians)	Evaluate pain (twice a day for 30 days)	Satisfaction	System Usability Scale (SUS)	All participants (100%) found the app extremely easy to use.	The total score of SUS is not presented
Spyridonis_2012 [42]	PainDroid	Assessment	End users N = 7 (wheelchair users over 18 years old)	1. Start the app 2. Rotate left/right, rotate up/down 3. Re-center 4. Zoom in/out and drag 5. Reset 6. Use above to select pain type and pain location on model 7. Save and exit	Satisfaction	Questionnaire adapted from SUS	The wheelchair users' evaluation highlighted positive bias in respect of the application's usability and functionality There was a positive trend in the question 3, referring to ease of use (Mean=4 St. Dev.=1)	Total score of the SUS adapted questionnaire is not presented

Vanderboom_2014 [43]	My Pain Diary	Assessment	End users N = 20 (patients who had complete an existing fibromyalgia treatment)	Enter the app to rate pain, fatigue and activity.	Satisfaction	1) Perceived Ease of Use (PEU) scale 2) Perceived Usefulness (PU) scale 3) Frequency of use	80% of the participants found the device easy to use, 75% found it easy to interact with and 70% thought it was easy to configure 75% of the users felt the app was useful to manage their condition Frequency of use over 7 days (in days) – 5.2±2.2(M±SD)
Bedson, J_2019 [44]	Keele Pain Recorder (KPR)	Assessment	End Users N= 21 (13 females and 8 males); median age 62 (IQR 50 to 70) Experts N = 17 (General Practitioners, physiotherapists, research nurses, primary care researchers, research facilitators, and an IT manager)	Enter pain recordings (twice daily for a period of 4 weeks)	Effectiveness Satisfaction	Completion rates Follow-up questionnaire Think aloud approach (Workshop) Semi-structured telephone interviews	Participants entered 862 records. Frequency of app use - 9/18 daily/often, 0/18 never. Patients found the KPR easy to use. General Practitioners found the graphical output easy to interpret.
Stefke_2018 [45]	Migraine Monitor	Assessment	End users N = 21 subjects (71% female, age 31±13 (M±SD))	Use the migraine diary function over a period of two weeks to register migraine attacks and to document daily mood changes.	Satisfaction	SUS	Overall, the application reached a SUS score of 91.67±6.96 (M ± SD). The average responses to the question "I think that I would use this system frequently" were considerably worse (2.71±1.41) compared to the average score of all other questions (3.77±0.52).
Turner-Bowker_2011 [46]	HEADACHE-CAT	Assessment	End users N = 9 headache sufferers (6 women and 3 men, age range, 27-54; mean, 38 years.	Participants navigated the survey and responded to the	Satisfaction	1) Structured interviews to gather user feedback on interface, visual	Participants reported that the stylus was ''easy'' or ''extremely easy'' to use.

				survey questions.		design and navigation; 2) Observation of the interviews and the interaction of users with the app; 3) Video recording of the process.	Participants were easily able to determine how to begin the survey again, close the application, return to a previous question, and proceed to a subsequent question.	
Huguet_2015 [47]	myWHI diary (my Wireless Headache Intervention diary)	Assessment	End users Cycle 1 (N=11) Cycle 2 (N=19) Cycle 3 (N=13) Characteristics Adolescents and young adults, age range 14-28years old.	Use the diary for 14 days. Report occurrence of headache: start and ending time, intensity, location, trigger(s), headache quality, highest pain intensity. Enter additional information (e.g., overall mood, hours and quality of sleep).	Effectiveness Satisfaction	Closed-ended questionnaire Adherence and completion rates Semi-structured interviews in the end of app use	A minority of participants completed all daily entries (18%, n=2) in cycle 1. In cycle 2, participants completed headache entries closer to the pain beginning time than in Cycle 1 (Cycle 2 mdn=3.83h vs Cycle 1 mdn=13.59h). Adherence was also improved, 26% of participants (n=5) completed all 14 daily entries. Cycle 3, the majority of participants' headache entries during Cycle 3 were made on the same day that the episode occurred. In all three cycles, most participants rated the app as easy to use and navigate.	
Minen_2018 [48]	RELAXaHEAD	Intervention	End users N = 12 Characteristics Adults with headache, 80% female (mean age was 37.5615.97, range 20–74 years).	Entry of daily headache diary data; Enter medications used; Register sleep data	Satisfaction	Think Aloud approach and Scripted protocol Recording of the sessions Likert-scaled questions about satisfaction	75% of the participants thought the diary was easy to use, 75% agreed that it was relevant to help track their headaches, 83% thought the information in the app was easy to understand, 75% agreed that the app kept their interest and attention during the session.	

							92% of the participants indicated that they would use the app again.	
Stinson_2013 [49]	Pain Squad	Assessment	End users Low-fidelity N = 15 adolescents (5 per cycle) High-fidelity N = 18 (1st cycle) N = 8 (2nd cycle) Experts N = 20 Characteristics 10 pediatric oncologists 10 pediatric pain experts	To complete the app, recording their own pain.	Satisfaction Effectiveness	Semi-structured, audio-taped interviews Register of entries in the system	Low-Fidelity Participants made positive comments on the appearance of the app screenshots; Participants suggested several changes to improve usability of the game-based app High Fidelity Every adolescent interviewed said they would use the app daily for an extended period; All participants found the app to be "easy to understand" and "easy to navigate". 100% entries were completed within the 2-week period.	
Hochstenbach_2016 [50]	No reference to app name	Intervention	End users N = 11 (patients with moderate to severe cancer pain, 5 males, 6 females)	1) Register pain, adverse effects, interference of pain with activity or sleep, and satisfaction with pain treatment; 2) register pain medication taken.	Satisfaction Effectiveness	Questionnaire to assess learnability, usability and desirability (Likert-scale) Completion rates Semi-structured interview	Patient quickly learned how to use the app (mean=4.8, SD=0.4). Average completion rates were 76.8% for pain monitoring, 50.4% for medication monitoring and 100% for education sessions. Patients completed, on average, 43 (SD 12.4) out of the 56 diaries. Overall, patient considered the app clear and simple.	
Fortier_2016 [51]	Pain Buddy	Intervention	End users N = 12 Characteristics Children with diagnose of cancer, ages 8-18 (12.33+/-3.42; 58% male)	Complete the symptom diary twice dailly for a 10 day period	Satisfaction Effectiveness	Queries with a likert-type scale about usefulness and ease of use;	Mean number of completed diaries – 19.58 (SD=4.64; range 13-27) Participants were satisfied with the app	

						Open-ended questions Completion rates of the diaries	
Neubert_2018 [52]	No reference to app name	Assessment	End users Study 1 N = 52 patients with chronic pain Study 2 N = 51 patients with chronic pain and N = 4 doctors	1) Make a correct symptom drawing on each view of the body outline showing the respective body region, 2) Specify the symptom choosing descriptors from a list, 3) Specify any pain-related symptom, 4) Rate the intensity of pain.	Satisfaction	**End users:** usability questionnaire, with 8 Likert-type questions and 2 open questions. **Experts:** Web-based survey comprising the questionnaires: 1) SUS 2) Attrakdiff 2 3) The ISONORM 9241/10 questionnaire.	End users In study 1, 87% (45/52) of the study patients were content with the body outline Experts Score of 75.63 on the SUS indicating an overall good usability
Fledderus_2015 [54]	No reference to app name	Intervention	End users N = 1 Characteristics Patient with chronic pain Experts N = 5 Characteristics 1 target group expert and 4 eHealth experts (mean age - 38, SD 12)	Scenarios with tasks/problems to solve (Tasks are not specified)	Satisfaction	Think Aloud approach (verbalizing thoughts during the scenarios) Audio recording of the sessions Evaluation of comments as positive, negative or neutral.	Ease of use of the app (e.g., clear navigation, clear buttons, simple) and the design (e.g., fresh, calm) were rated positively. Usefulness of the features also had positive comments. Most negative comments were about the quality of the system as some (technical errors and unclear icons).
Cardos_2017 [55]	BIO-H App	Intervention	End users N = 16	1) opening the intervention app;	Effectiveness Satisfaction	System Usability Scale	A mean score of 93.44 (SD=7.06) out of 100 was found for the

			Characteristics Females (mean age = 42,43)	2) read the information about the intervention; 3) fill pre-intervention survey; 4) listen to the audio relaxation exercise; 5) fill out the post-intervention survey		Structured Questionnaire Task completion rates	overall usability based on the participants' SUS ratings. People found the intervention component to be useful (m=8.63 out of 10; SD = 1.14) and they indicated a high level of willingness to access the app in the future (m=8.56; out of 10; SD=1.71). All tasks were completed successfully, and participants commented on how easy the interface was to navigate.	
Boceta, J_2019 [56]	App INES·DIO	Assessment	End users N = 175 physicians	Open a new profile for each patient and go through four sequential steps to collect relevant clinical information for an enhanced diagnosis procedure of BTcP.	Satisfaction Effectiveness	Structured questionnaire Assessment test related to usability (included as last feature of the app) Follow-up questions about the app design, overall usefulness, and use intention Utilization rates of the app and each assessment scale	Most clinicians (157/175, 89.7%) concluded that the mobile app is well designed and easy to use 94.9% (166/175) of participants would likely/most likely recommend the use of App INES·DIO. The app was used as often as two or three times a week by 41.7% (74/175) of clinicians. Pain-rating and functional assessment scales were the most frequent tools used by clinicians (93.3%, 163/175)	Boceta, J_2019 [56]
Docking_2018 [57]	iPhone pain app	Assessment	End users N = 24 Characteristics Trainee paramedics (46% females and 54% males) Experts	Participants were asked to work their way through the app, trialling out different routes.	Satisfaction	4 focus groups with trainee paramedics Direct questions related to changes/ recommendations	Focus group revelled concerns around the usability of technological devices within the paramedic setting and whether it is professional to use them in practice.	

| | | | N= 7
Characteristics
Paramedics (86% males
and 14% females) | | | Delphi Panel of
Paramedic Experts | Some suggestions were related
to making the app more
intuitive, in particular the
location of the "next" and
"back" buttons and the
correction of some links. | |