Clinical Decision Support for Parents through Mobile Applications: A Systematic Assessment of Pediatric Fever Management Apps

Chloe Joosen [a], Jaan Toelen [b,c,d], Willeke Asscherickx [e]

[a] Faculty of Medicine, KU Leuven, Belgium
[b] Leuven Child and Youth Institute, KU Leuven, Belgium
[c] Department of Paediatrics, University Hospitals Leuven, Belgium
[d] Department of Development and Regeneration, KU Leuven, Belgium
[e] Department of Paediatrics, AZ Diest, Belgium

w.asscherickx@azdiest.be

Keywords

Medical decision tools ; mobile applications ; triage ; fever.

Abstract

Objective

This study aimed to identify existing apps for paediatric fever management and compare their decision algorithms with current evidence-based guidelines.

Methods

From May to July 2022, mobile applications were systematically searched in the Apple App Store and Google Play Store using specific terms. The apps underwent four rounds of screening to match predefined criteria. Each app was evaluated by five independent reviewers using the Mobile Application Rating Scale (MARS). The decision support algorithms of each app were analysed for adherence to existing fever management recommendations, including the NICE guidelines for children under five years of age and the Schmitt-Thompson triage protocol for children under and over three months of age.

Results

Out of 878 apps retrieved, 6 met the selection criteria and were 3 finally assessed. The apps scored high on overall quality, averaging a MARS rating of 4.3 out of 5. Kinsa and FeverApp scored the highest (4.4 out of 5), followed by FeverFriend (4.0 out of 5). FeverFriend showed the highest adherence to the NICE guidelines, followed by Kinsa and FeverApp. For the Schmitt-Thompson protocol, Kinsa showed the highest adherence, followed by FeverFriend and FeverApp.

Conclusion

The availability of evidence-based fever management apps with parental decision support systems is limited. Kinsa emerged as the top-performing app based on quality assessment.

Introduction

Fever, defined as a body temperature exceeding the normal range (36.6°C to 37.9°C), is a natural physiological response that boosts the immune system and combats infection, potentially shortening the duration of illness (1). It is a common childhood condition as children between the ages of 3 and 36 months typically experience about six febrile episodes annually (2). This underscores the fact that fever is usually a symptom of other illnesses rather than a disease itself (3). Although fever in children under five is often linked to benign viral infections, it can also be associated with serious health issues.

This makes fever management very challenging for parents, mostly due to misconceptions about its significance and effects (4). Caregivers fear harm either from the fever itself or the underlying illness (5). They are thus confronted with the decision to treat symptoms at home or seek medical attention, which leads to overuse primary care and specialized secondary care systems (6,7).

The surge in mobile technology usage has led to the emergence of health-related applications, including those aimed at assisting parental decision-making. These apps, which include symptom checkers and self-triage tools, promise to be transformative, yet their integration of evidence-based medicine remains inadequately examined (8, 9, 10). This study focuses on evaluating mobile applications equipped with parental decision support systems (PDSS) for the management of paediatric fever. We aim to compare these apps' decision-making algorithms with current evidence-based guidelines and examine their potential in aiding caregivers to make informed choices about managing fever, which could help reduce unnecessary use of antipyretics and reduce the burden on healthcare facilities.

Material and Methods

Search Strategy

The initial identification of existing triage applications, was obtained by entering the keywords "pediatrics" OR "fever" OR "pediatrics, fever" OR "pediatrics, symptom" OR "fever, children" OR "fever, child" OR "fever, symptom" OR "febrile, symptom" in the search engines of both the Apple App store for iOS and the Google Play store for Android. Searches were performed within these digital distribution platforms from May to July 2022.

After the identification process, preliminary selection and assessment of mobile applications was based upon a prespecified set of inclusion and exclusion criteria as defined in Table 1. Selected applications were subject to four rounds of screening.

Since the apps were obtained by systematic searches in two sources (the Apple App Store and the Google Play Marketplace) using multiple search terms, results were screened for duplicates. Duplicate apps, i.e. identical features in the same app listed independently in both app stores or listed in the same app store using different search terms, were excluded in the first round. Applications by the same developer, similarly named, but not having identical feature sets were treated as two individual apps.

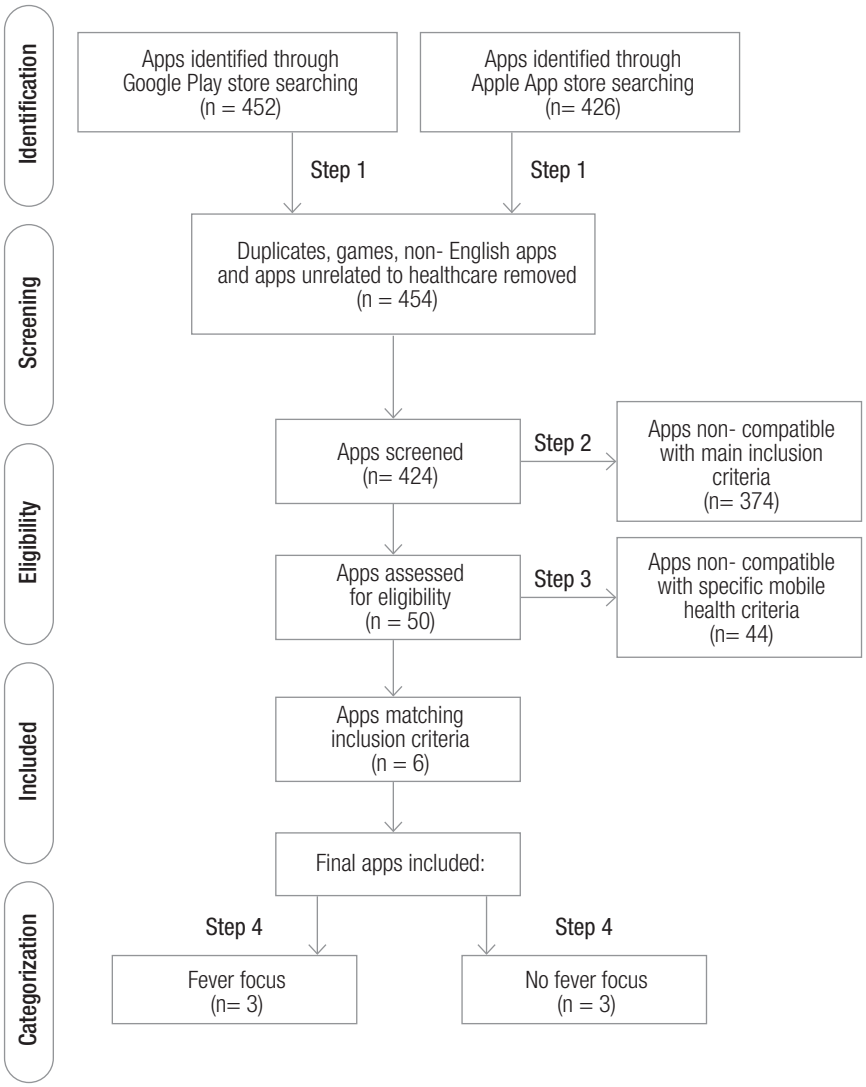
Table 1: Inclusion and exclusion criteria.

Main inclusion criteria	Specific mobile health criteria	Exclusion criteria
• Children health assessment • Pediatrics • Fever management support	• Parental decision support system • No exclusive focus on fever monitoring and fever registration • No exclusive focus on telemedicine/ video consultation • Ability to function independently of a medical device • Ability to function without possession of a US phone number	• Games • No relation to healthcare and medical services • No English user interface • Health care workers as primary audience

Additionally, games or applications without presence of English user interface were excluded in the first round as well. The second round of screening was performed based on title/name and available summary description within the app stores. Applications that were non- compatible with main inclusion criteria were excluded. The general target population in this study includes non- clinical app users, i.e. parents and caregivers. Apps aimed at the provision of fever management guidance within clinical settings, that is with target users being physicians and healthcare professionals, were also excluded as a part of round two.

In the third round, the remaining applications were downloaded and manually assessed for evaluation of specific mobile health criteria. The apps inconsistent with our criteria were eliminated. In the fourth round apps with main focus on fever management in children were differentiated from those that provide fever management decision support, without fever management in children being the app's main target, by dividing them into two groups. Thus, the first group of apps focuses entirely on assisting parents who are dealing with a febrile child while the second group of apps provide a decision support system for numerous symptoms in a variety of ages, including fever. The flowchart in Figure 1 illustrates the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology described.

Figure 1: Flowchart of the selection process.



Quality assessment: the Mobile Application Rating Scale

First, the applications were independently evaluated by a team of five reviewers (parents without medical background) using the Mobile Application Rating Scale (MARS). MARS is a widely used standardized tool developed by the Queensland University of Technology. It provides a multidimensional measure of mobile health app quality by assessment in three sections. The first section is divided into four objective dimensions, i.e. engagement, functionality, aesthetics, and information quality. In short, the scale measures whether an app is interesting, interactive (e.g. sends alerts, messages, reminders or feedback) and easy to learn. It analyses the gestural design and examines the quality and quantity of information provided. In the first section nineteen items are rated on a 5-point scale (1= inadequate, 2= poor, 3= acceptable, 4= good, and 5= excellent). The mean score of each dimension is then calculated and used to compute the app overall quality score (11, 12). The second and third section, being the app subjective quality score and the app specific quality score, are equally determined by rating of respectively 4 and 6 additional items. Supplementary, ambiguous items were discussed and clarified to ensure full comprehension of the scale (13).

Table 2: Adherence of existing fever management apps to the NICE guideline concerning fever under the age of five.

NICE guideline	FeverFriend	Kinsa	FeverApp*	Points of comparison
FEVER				
Fever (T < 38°C) Under 3 months	No Fever (T ≤ 38°C) Under 3 months	No Fever (T < 38°C) Under 3 months	No Fever (T ≤ 38°C) Under 4 months	**Fever under 3 months**: Both the Feverfriend and Kinsa app identified the symptom as high risk, thereby matching the guideline. The FeverApp app underestimated the symptom compared to the guideline, by classifying it as an intermediate risk of serious illness. The FeverApp app additionally mentioned both 3 months and 4 months as a cut off in age. **High Fever between 3 to 6 months**: None of the apps matched the guideline perfectly. Both the Feverfriend and Kinsa app identified 'fever (T ≥ 38°C) between 3 and 6 months' as an intermediate risk of serious illness, thereby overestimating the symptom compared to the guideline. The FeverFriend app additionally identified a 'high fever (T ≥ 39°C)' within this age group as a high risk symptom.The FeverApp app did not mention this symptom. **Fever duration**: None of the apps matched the guideline perfectly. The FeverFriend app identified a 'fever duration > 5 days' as a high risk symptom, thereby overestimating the symptom compared to the guideline. The FeverApp app used a different cut off of in time of 3 days. The Kinsa app did not mention this symptom. **Type of measurement**: Every app recommended fever measurement by usage of a infra-red tympanic thermometer, thereby matching the guideline. Yet, placement preference and usage of the electronic and chemical dot thermometer varied among the apps and did not match the guideline. **Rigors**: The Kinsa app underestimated the symptom compared to the guideline by classifying it as a low risk of serious illness.The FeverFriend and FeverApp app did not mention the symptom.
Fever (T < 39°C) Between 3-6 months	No Fever (T ≤ 38°C) Between 3-6 months	No Fever (T < 38°C) Between 3-6 months		
Fever (T ≥ 38°C) Above 6 months	No Fever (T ≤ 40°C) Above 6 months	No Fever (T < 40°C) Above 6 months	Fever (T > 38°C) Above 4 months	
Electronic thermometer axillary Chemical dot thermometer axillary Infra- red tympanic thermometer	Digital thermometer rectal/oral * Chemical thermometer rectal/oral * Infra- red tympanic thermometer	Rectal- type of thermometer unspecified * Infra- red tympanic thermometer *	Rectal - type of thermometer unspecified * Infra- red tympanic thermometer	
Fever < 5 days	Fever < 3 days		Fever < 3 days	
No rigors		Chills		
High Fever (T ≥ 39°C) Between 3-6 months	Fever (T > 38°C) Between 3-6 months	Fever (T ≥ 38°C) Between 3-6 months	Fever (T > 38°C) Under 4 months	
	High Fever (T > 40°C) Above 6 months			
Fever ≥ 5 days	Fever 3-5 days		Fever > 3 days	
Rigors				
Fever (T ≥ 38°C) Under 3 months	Fever (T > 38°C) Under 3 months	Fever (T ≥ 38°C) Under 3 months		
	High Fever (T > 39°C) Between 3-6 months			
	Very high Fever (T > 41°C) Above 6 months	High Fever (T ≥ 40°C)		
	Fever > 5 days			
SKIN CONDITION AND COLOR				
Normal color of skin, lips and tongue	Normal color of skin	Normal color of skin and lips		**Color of skin**: Both the FeverFriend and Kinsa app identified a bluish skin as a high risk symptom, thereby matching the guideline. However, the FeverFriend app matches the guideline in its entirety by also classifying pallor as an intermediate risk symptom. The FeverApp app did not include skin color in its triage system. **Rash**: Both the Feverfriend and the FeverApp app identified a non-blaching rash as a high risk symptom, thereby matching the guideline. However, the FeverFriend app matches the guideline in its entirety by also classifying a blaching rash as an intermediate risk symptom. The FeverApp app identified a blaching rash as an intermediate risk symptom, only when present for more than 3 days, thereby not perfectly matching the guideline. The kinsa App specified the type of rash in its triage system, where the differentiation between a blanching rash as an intermediate risk symptom and a non- blaching rash as a high risk symptom remained unclear.
No rash	No rash	No rash	No rash	
		Rash Above 2 years	Blanching rash Present for < 3 days	
Pallor	Pallor			
Blanching rash	Blanching rash	Rash Under 2 years	Blanching rash Present for > 3 days	
		Hives		
Pale, mottled, ashen, blue skin	Grey, bluish, purplish skin	Bluish skin or lips		
Non- blanching rash	Non- blanching rash	New purple or blood- colored spots/dots	Non- blanching rash	
		Blisters, sores or pus		
		Rash developing after taking new medication		
		Sudden and severe skin peeling		
RESPIRATORY				
No tachypnea *	No tachypnea *	No tachypnea *	No tachypnea *	**Tachypnea**: The FeverFriend app did not match the guideline perfectly by having somewhat different cut offs to differentiate tachypnea and severe tachypnea and classifying them as a intermediate and high risk symptom, respectively. The Kinsa app identified a respiratory rate above 60 breaths per minute as a high risk symptom, thereby matching the guideline. No further specifications were made. The FeverApp app defined a respiratory rate above 60 breaths per minute as an intermediate risk symptom, thereby underestimating the symptom compared to the guideline. Breathing frequency cut offs are specified below. **Signs of dyspnea**: The Kinsa and FeverApp app both defined shortness of breath by using a series of clinical observations. The Kinsa app matched few clinical observations compared to the guideline by classifying 'grunting' and the presence of 'retractions' or 'chest indrawing' as a high risk symptom. 'Nasal flaring' was classified as a high risk symptom, thereby overestimating the symptom compared to the guideline.The FeverApp app matched no clinical observations to the guideline. The FeverFriend app defined shortness of breath trough a scale, making comparison to the guideline difficult. The scale is specified below.
No signs of dyspnea	No signs of dyspnea	No signs of dyspnea	No signs of dyspnea	
Tachypnea *	Tachypnea *		Severe tachypnea *	
Nasal aring				
	Difficult or laboured breat hing ≤ 3/5 *		Enforced or constrained breathing	
	Slight wheezing			
		Barking cough		
Severe tachypnea *	Severe tachypnea *	Severe tachypnea *		
		Difficult breathing		
		Unable to finish a sentence		
		Nasal flaring		
Grunting		Grunting		
Moderate or severe chest indrawing		Retractions		
	Strong wheezing, stridor	Wheezing, stridor		
CIRCULATION AND HYDRATION				
No tachycardia *	No tachycardia *		No tachycardia *	**Tachycardia**: None of the apps matched the guideline perfectly.The FeverFriend and FeverApp app both defined anomalous cut offs compared to the guideline when defining tachycardia. The FeverFriend app differentiated tachycardia and severe tachycardia whereas the guideline did not. The Kinsa app did not mention this symptom. Heart rythm cut offs are specified below. **Poor feeding in infants**: The FeverApp app classified the symptom 'last time drinking > 8 hours in infants' as an intermediate risk symptom, thereby matching the guideline. The guideline does not specify the symptom in time, therefore 'last time drinking > 8 hours in infants' was considered a match to the guideline. The FeverFriend app differentiated poor and severly poor feeding by cut off in time. However, since no specification for infants was made by the app, as stated within the guideline, the symptom did not match the guideline perfectly. The Kinsa app equally did not specify an age group and identified 'last time drinking > 8 hours' as a high risk symptom, thereby not matching the guideline. **Dry mucous membrane**: The FeverApp app identified the symptom as an intermediate risk symptom, thereby matching the guideline. The Kinsa app overestimated the symptom compared to the guideline by classifying it as a high risk symptom. The FeverFriend app differentiated a 'dry tongue with decreased tears' from 'absence of tears' by classifying it as an intermediate and high risk symptom, respectively. Since the guideline mentions 'absence of tears' as part of 'dry mucous membrane', classifying 'no tears when c rying' as a high risk symptom overestimates the guideline.Therefore, the FeverFriend app only matched the guideline partially. **Reduced skin turgor**: The FeverFriend app differentiated a somewhat reduced skin turgor from a severely reduced skin turgor by classifying it as an intermediate risk symptom and a high risk symptom, respectively. However since the guideline defines any form of reduced skin turgor as a high risk of serious illness, the symptom did not match the guideline perfectly. The FeverApp app underestimated the symptom compared to the guideline, by classifying it as a intermediate risk symptom. The Kinsa app dit not mention this symptom. **Reduced urine output**: The FeverFriend app differentiated a reduced urine output from a severely reduced urine output by classifying it as an intermediate risk symptom and a high risk symptom, respectively. The guideline does not specify the symptom in time, therefore 'last urination > 6-12 hours' was considered a match to the guideline. The FeverApp app matched the guideline by classifying the symptom as an intermediate risk symptom. The Kinsa app overestimated the symptom compared to the guideline by classifying it as a high risk symptom
No signs of dehydration	No signs of dehydration	No signs of dehydration	No signs of dehydration	
Tachycardia *	Tachycardia *		Tachycardia *	
Poor feeding in infants	Drinking less than normal Last time eating > 12 hours		Last time drinking > 8 hours in infants *	
Dry mucous membrane *	Dry tongue, decreased tears when crying		Dry mucous membrane *	
CRT ≥ 3 seconds				
	Somewhat reduced skin turgor		Reduced skin turgor	
	Reduced urine output (Last urination > 6-12 hours)		Reduced urine output (Not specified)	
	Severe tachycardia *			
	Last time drinking > 12 hours Last time eating > 24 hours		Last time drinking > 8 hours	
	No tears when crying		Dry mucous membrane *	
Reduced skin turgor	Severely reduced skin turgor			
	Severely reduced urine output (Last urination > 12 hours)		Reduced urine output (Last urination > 8 hours)	
GENERAL CONDITION AND ACTIVITY				
Normal response to social cues	Normal response to social cues	Normal reactions to social cues, making eye contact	Normal response to social cues	**Response to social cues**: A normal respons to social cues is not specified in content or time by the guideline. The FeverFriend app defined 'odd reaction > 5 hours' as an intermediate risk symptom, however since the lack of specificity, it was considered no match to the guideline. Both the FeverFriend and the Kinsa app mentioned clinical observations indicating an absence of response to social cues. These symptoms were classified as high risk symptoms, thereby matching the guideline. The FeverApp app defined 'acting differently' and 'apathy' as high risk symptoms. However since the lack of specificity, comparison to the guideline remained difficult and no match could be concluded. **Awareness**: All apps identified an increased tendency to fall asleep as an intermediate risk symptom, thereby matching the guideline by indicating a slightly decreased state of awareness and a decreased level of acitivity. A severely altered conciousness was classified as a high risk symptom by all apps, thereby matching the guideline. **Irritability**: No app identified the absence of smile as an intermediate risk symptom. The Kinsa app identified a crooked smile as a high risk symptom, thereby not matching the guideline. All the apps identified a continous, weak or high pitched cry or screaming (as a sign ofirritability) as a high risk symptom, thereby matching the guideline. **Swelling or pain of joint or limb**: The FeverApp app identified 'swelling of a joint' as an intermediate risk symptom, thereby matching the guideline. The FeverFriend app identified 'swelling of body part' as a high risk symptom, thereby overestimating it compared to the guideline. Avoidance of using an extremity was not mentioned by any of the apps. Howerever, the FeverFriend app did identify 'protection of body part' as a high risk symptom. **Seizure**: The guideline defines 'status epilepticus' or 'a focal seizure' as high risk symptoms. The Kinsa and the FeverApp app both identify status epilepticus as a high risk symptom, thereby matching the guideline. A feverish seizure itself was equally classified by all apps as a high risk symptom. The FeverApp app specified only the first seizure as a high risk symptom. Ambiguity remains within the guideline concerning the management of a feverish seizure. **Bulging fontanelle**: The FeverFriend app underestimated the symptom compared to the guideline, by classifying it as an intermediate risk symptom. The Kinsa app identified the symptom as a high risk symptom, thereby matching the guideline. The FeverApp app did not mention this symptom. **Stiff neck**: The FeverFriend and the Kinsa app both identified the symptom as a high risk symptom, thereby matching the guideline. The FeverApp app underestimated the symptom compared to the guideline, by classifying it as an intermediate risk symptom.
Able to stay awake or awakens quickly, no decreased activity	Normal awareness	Normal awareness	No lethargy	
No signs of irritability or pain; content or smiles, not crying or strong crying	No signs of irritability or pain; not crying or strong crying	No signs of irritability or pain	No signs of irritability or pain	
No swelling of a limb or joint and normal usage of all extremities	No swelling of a body part and normal usage of all extr emities		No swelling of the joints	
No seizure or focal neurological signs	No seizure	No seizure	No seizure	
No bulging fontanelle	No bulging fontanelle	No bulging fontanelle		
No stiff neck	No stiff neck	No stiff neck	No stiff neck	
Not responding normally to social cues	Odd reactions > 5 hours			
Wakes only with prolonged stimulation, decreased activity	Sleepy	Fatigue	Lethargy	
No smile				
Swelling of a limb or joint			Swelling of a joint	
Non- weight bearing/ not using an extremity			Pain in the limb ≥ 3 days	
	Bulging fontanelle			
			Stiff neck	
No response to social cues	No reaction to social cues	Not making eye contact, acts 'out of it' or confused	Acting differently, apathy	
Unable to rouse or if roused does not stay awake	No awareness	Inability to wake up, unresponsive	Clouded consciousness	
Weak, high- pitched or continuous cry	Weak, high- pitched or continuous cry	Inconsolable crying, shrill screaming	Shrill screaming	
		Touch sensitivity	Touch sensitivity	
		Crooked smile		
	Seizure	Seizure	First seizure	
Status epilepticus		Status epilepticus	Status epilepticus	
Focal seizures				
Focal neurological signs				
Bulging fontanelle		Bulging fontanelle		
Stiff neck	Stiff neck	Stiff neck		
		Swelling of body part(s)		
		Protection of body part(s)		

Table 3: Adherence of existing fever management apps to the Smitt- Thompson protocol concerning fever in children < and > 3 months of age.

Smitt- Thompson protocol	Kinsa	FeverFriend	FeverApp*	Points of comparison
FEVER				
No fever (T ≤ 38°C) Under 3 months	**No Fever (T < 38°C) Under 3 months**	**No Fever (T ≤ 38°C) Under 3 months**	**No Fever (T ≤ 38°C) Under 4 months**	**Fever under 3 months**: Both the Feverfriend and Kinsa app identified the symptom as high risk, thereby matching the guideline. The FeverApp app underestimated the symptom compared to the guideline, by classifying it as an intermediate risk of serious illness. The FeverApp app additionally mentioned both 3 months and 4 months as a cut off in age. **High Fever between 3 to 6 months**: None of the apps matched the guideline perfectly. Both the Feverfriend and Kinsa app identified fever (T ≥ 38°C) between 3 and 6 months as an intermediate risk of serious illness, thereby overestimating the symptom compared to the guideline. The FeverFriend app additionally identified a high fever (T ≥ 39°C) within this age group as a high risk symptom.The FeverApp app did not mention this symptom. **Very high fever**: The guideline identified a cut off of 40.6 °C body termperature. Both the FeverFriend and Kinsa app specified a different cut off, thereby not matching the guideline. The FeverApp app did not mention this symptom. **Fever duration**: The FeverFriend and the FeverApp app both identified a 'fever duration > 3 days' as an intermediate risk symptom, thereby matching the guideline. However, the FeverFriend app mentioned another cut off in time of 5 days, classifying a 'fever duration > 5 days' as a high risk symptom and thereby not matching the guideline in its enterity. The Kinsa app did not mention this symptom. **Type of measurement**: Every app recommended fever measurement rectally, thereby matching the guideline. However, the guideline specifies the age from wich oral en tympanic fever measurements are recommended. Due to absent age specification by the apps, no additional match to the guideline could be concluded. **Rigors**: The Kinsa app did not specify a time lapse within the presence of chills and identified the 'presence of chills' as a low risk symptom, thereby underestimating the symptom compared to the guideline.The FeverFriend and FeverApp app did not mention the symptom.
Fever (T ≤ 39°C) Between 3-6 months	No Fever (T < 38°C) Between 3-6 months	No Fever (T ≤ 38°C) Between 3-6 months		
Fever (T ≤ 40.6°C) Above 6 months	No Fever (T < 40°C) Above 6 months	No Fever (T ≤ 40°C) Above 6 months	Fever (T > 38°C) Above 4 months	
Rectal - type of thermometer unspecified Oral - type of thermometer unspecified (not under 3 months) Infra- red temporal artery thermometer Inra- red tympanic thermometer (not under 6 months)	**Rectal- type of thermometer unspecified * **Infra- red tympanic thermometer * - age unspecified	**Rectal - digital or chemical thermometer *** Oral - digital or chemical thermometer - age unspecified * Infra- red tympanic thermometer - age unspecified	**Rectal - type of thermometer unspecified *** Infra- red tympanic thermometer - age unspecified	
Fever < 3 days		**Fever < 3 days**	**Fever < 3 days**	
No shaking chills	Chills			
High Fever (T ≥ 39°C) 3-6 months	Fever (T ≥ 38°C) Between 3-6 months	Fever (T > 38°C) Between 3-6 months	Fever (T > 38°C) Under 4 months	
	High Fever (T > 40°C) Above 6 months			
Fever > 3 days		**Fever 3-5 days**	**Fever > 3 days**	
Fever (T ≤ 38°C) Under 3 months	**Fever (T ≥ 38°C) Under 3 months**	**Fever (T > 38°C) Under 3 months**		
Very high Fever (T ≥ 40.6°C)	High Fever (T ≥ 40°C)	High Fever (T > 39°C) Between 3-6 months		
		Very high Fever (T > 41°C)		
		Fever > 5 days		
Shaking chills > 30 minutes				
SKIN CONDITION AND COLOR				
Normal color of skin, lips and tongue	**Normal color of skin and lips**	**Normal color of skin**		**Color of skin**: Both the FeverFriend and Kinsa app identified a bluish skin as a high risk symptom, thereby matching the guideline. The FeverApp app did not include skin color in its triage system. **Rash**: The Kinsa app identified the presence of 'purple or blood- colored spots' as a high risk symptom, thereby matching the guideline. The FeverFriend app and the FeverApp app did not specify a 'non-blanching rash'. Due to absent content specification by the apps, no match to the guideline could be concluded.
No rash	**No rash**	**No rash**	**No rash**	
	Rash Above 2 years		Blanching rash Present for < 3 days	
		Pallor		
	Rash Under 2 years	Blanching rash	Blanching rash Present for > 3 days	
	Hives			
Bluish lips or face	**Bluish skin or lips**	**Grey, bluish, purplish skin**		
Widespread rash with purple or blood- colored spots or dots	**New purple or blood- colored spots/dots**	Non- blanching rash	Non- blanching rash	
	Blisters, sores or pus			
	Rash developing after taking new medication			
	Sudden and severe skin peeling			
RESPIRATORY				
No signs of dyspnea	**No signs of dyspnea**	**No signs of dyspnea**	**No signs of dyspnea**	**Signs of dyspnea**: The Kinsa and FeverApp app both defined shortness of breath by using a series of clinical observations. The Kinsa app matched some clinical observations compared to the guideline by classifying 'grunting' and being 'unable to finish a sentence' as a high risk symptom. Both the FeverFriend and the Kinsa app identified difficult breathing as a high risk symptom, thereby matching the guideline. However, the FeverFriend app defined shortness of breath trough a scale, thereby making comparison to the guideline difficult. The '> 4/5' was concluded as similar to 'struggling for each breath', therefore a partial match to the guideline was concluded.The scale is specified below. The FeverApp app matched no clinical observations to the guideline. **Excessive spitting, drooling**: The Kinsa app identified the symptom as a high risk symptom, thereby matching the guideline. The FeverFriend and FeverApp app did not mention this symptom.
No spitting of saliva	**No spitting of saliva or fluids**			
		Difficult or laboured breathing ≤ 3/5 *	Enforced or constrained breathing	
		Slight wheezing		
	Barking cough			
	Nasal flaring			
Difficult breathing, struggling for each breath	**Difficult breathing**	**Difficult or laboured breathing > 4/5 ***		
Unable to speak or cry	**Unable to finish a sentence**			
Grunting	**Grunting**			
	Retractions			
	Wheezing, stridor	Strong wheezing, stridor		
Unable to swallow fluid or excessive spitting	**Unable to swallow fluid or excessive spitting**			
CIRCULATION AND HYDRATION				
No signs of dehydration	**No signs of dehydration**	**No signs of dehydration**	**No signs of dehydration**	**Poor feeding in infants**: The FeverFriend app differentiated poor and severly poor feeding by cut off in time. However, since no specification for infants was made by the app, as stated within the guideline, the symptom did not match the guideline perfectly. The FeverApp app classified the symptom 'last time drinking > 8 hours in infants' as an intermediate risk symptom, thereby underestimating the symptom compared to the guideline. The Kinsa app equally did not specify an age group and identified 'last time drinking > 8 hours', thereby not matching the guideline perfectly. D**ry mucous membrane**: The Kinsa app identified the symptom as a high risk symptom, thereby matching the guideline.The FeverFriend app differentiated a 'dry tongue with decreased tears' from 'absence of tears' by classifying it as an intermediate and high risk symptom, respectively. Since the guideline mentions 'absence of tears' as part of 'dry mucous membrane', classifying 'no tears when crying' as a high risk symptom matched the guideline. However since a dry tongue was identified as an intermediate risk symptom and the guideline mentions 'dry mouth' as part of 'dry mucous membrane', the FeverFriend app only matched the guideline partially. The FeverApp app identified the symptom as an intermediate risk symptom, thereby not matching the guideline. **Reduced urine output**: The FeverFriend app differentiated a reduced urine output from a severely reduced urine output by classifying it as an intermediate risk symptom and a high risk symptom, respectively. The guideline does not specify the symptom in time, therefore 'last urination > 12 hours' was considered a match to the guideline. The Kinsa app identified the symptom as a high risk symptom, thereby matching the guideline. The FeverApp app identified the symptom as an intermediate risk symptom, thereby underestimating it compared to the guideline.
		Drinking less than normal Last time eating > 12 hours	Last time drinking > 8 hours in infants *	
		Dry tongue, decreased tears when crying	Dry mucous membrane *	
		Reduced urine output (Last urination > 6-12 hours)	Reduced urine output (Not specified)	
Poor feeding in infants	Last time drinking > 8 hours	Last time drinking > 12 hours Last time eating > 24 hours		
Dry mucous membrane *	**Dry mucous membrane ***	**No tears when crying ***		
Reduced urine output	**Reduced urine output (Last urination > 8 hours)**	Severely reduced urine output (Last urination > 12 hours)		
GENERAL CONDITION AND ACTIVITY				
No altered mental status or decreased activity	**Normal awareness**	**Normal awareness**	**No lethargy**	**Altered mental status**: An altered mental status was described by the guideline as a confused state with impaired alertness. Both the FeverFriend and the Kinsa app mentioned clinical observations indicating an altered mental status. These symptoms were classified as high risk symptoms, thereby matching the guideline. The FeverApp app defined 'acting differently' and 'apathy' as high risk symptoms. However, since the lack of specificity, comparison to the guideline remained difficult and no match could be concluded. The FeverFriend app defined 'odd reaction > 5 hours' as an intermediate risk symptom. No intermediate risk symptoms were specified by the guideline. **Awareness**: A limited conciousness or absent awareness was classified as a high risk symptom by all apps, thereby matching the guideline. All apps additionally identified an increased tendency to fall asleep as an intermediate risk symptom. No intermediate risk symptoms were specified by the guideline. **Irritability**: All the apps identified an inconsolable, weak or high pitched cry or screaming (as a sign of irritability) as a high risk symptom, thereby matching the guideline. The Kinsa and the Feverapp app both additionally identified 'touch sensitivity' as a high risk symptom, thereby matching the guideline in its entirety. **Usage of extremities**: Avoidance of using an extremity was not mentioned by any of the apps. Howerever, the FeverFriend app did identify 'protection of body part' as a high risk symptom. Since the guideline specified an extremity, 'protection of body part' was considered too broad to match the guideline perfectly. **Seizure**: The FeverFriend and Kinsa app both identified a feverish seizure as a high risk symptom, thereby matching the guideline. The FeverApp app specified only the first seizure as a high risk symptom, thereby not matching the guideline perfectly. **Bulging fontanelle**: The Kinsa app identified the symptom as a high risk symptom, thereby matching the guideline. The FeverFriend app underestimated the symptom compared to the guideline, by classifying it as an intermediate risk symptom. The FeverApp app did not mention this symptom. **Stiff neck**: The FeverFriend and the Kinsa app both identified the symptom as a high risk symptom, thereby matching the guideline. The FeverApp app underestimated the symptom compared to the guideline, by classifying it as an intermediate risk symptom. **Painfull urination**: The FeverFriend app identified the symptom as an intermediate risk symptom, thereby underestimating it compared to the guideline. The Kinsa and FeverApp app did not mention this symptom.
No signs or irritability or pain	**No signs of irritability or pain**	**No signs of irritability or pain; not crying or strong crying**	**No signs of irritability or pain**	
Normal usage of all extremities		No swelling or pain of a body part	No swelling or pain of the joints	
No seizure	**No seizure**	**No seizure**	No seizure	
No bulging fontanelle	**No bulging fontanelle**	**No bulging fontanelle**		
No stiff neck	**No stiff neck**	**No stiff neck**	**No stiff neck**	
No painful urination		**No painful urination**		
		Odd reactions > 5 hours		
	Fatigue	Sleepy	Lethargy	
			Swelling of a joint	
			Pain in the limb ≥ 3 days	
		Bulging fontanelle		
			Stiff neck	
		Painful urination		
Altered mental status; awake but not alert, not focused, confused or slow to respond	**Not making eye contact, acts 'out of it' or confused**	**No reaction to social cues**	Acting differently, apathy	
Unresponsive or difficult to waken	**Inability to wake up, unresponsive**	**No awareness**	**Clouded consciousness**	
Limp, weak or not moving	**Limp, not moving**			
Inconsolable crying, irritable	**Inconsolable crying, shrill screaming**	**Weak, high- pitched or continuous cry**	**Shrill screaming**	
Cries every time if touched, moved or held	**Touch sensitivity, crying when touched**		**Touch sensitivity**	
Seizure	**Seizure**	**Seizure**	First seizure	
Bulging fontanelle	**Bulging fontanelle**			
Stiff neck	**Stiff neck**	**Stiff neck**		
		Swelling of body part(s)		
Won't move arm or leg normally		Protection of body part(s)		
Burning feeling or pain with urination				

Statistical Analysis

The MARS scores were subjected to descriptive analysis. Interrater reliability (IRR) of the MARS subscales and total score was additionally determined by calculating the intraclass correlation coefficient (ICC). This statistic allows assessment of rater agreement and is a number found to have a value between 0 and 1. Zero indicating no reliability among raters and one indicating perfect reliability among raters. ICC estimates and their 95% confident intervals were calculated using SPSS statistical package version 28, based on a mean- rating (k=5), absolute- agreement, 2- way mixed- effects model (14).

Quality assessment: adherence to EBM guidelines

Secondly, adherence to existing fever management recommendations was appraised through comparison against the National Institute of Care and Excellence (NICE) guideline on fever in children under five years of age (15). The Schmitt-Thompson triage protocol for fever in children (under 3 months to 3 months to 6 years of age) was also used for evaluation (16). Our goal was to gain a better understanding of the decision algorithm being used in these mobile applications promoting fever management, as well as to review whether the propositions made are in line with existing prespecified evidence and recommendations. Therefore the mobile applications with a main focus on paediatric fever management were each manually run using 5 profiles with 5 different ages, i.e. 1 month, 3 months, 6 months, 1 year and 5 years. Within each age group, the decision support algorithm was mapped by exploration (given that each app contains a decision support system as stated in the specific mobile health inclusion criteria).

By mapping out these answer options, a set of symptoms was identified validating an alteration in advice. These disease symptoms were stratified into a traffic light system. For example, when the app adjusted its guidance to an advice expressing an urgent need for consultation, the present symptom upon which this adjustment was based was qualified within the colour red indicating a high risk of serious illness. When the app adjusted its guidance to an advice expressing a non-urgent need for consultation the present symptom upon which this adjustment was based, was qualified within the colour yellow indicating an intermediate risk of serious illness. When the app's advice was 'home care', expressing an absent need for consultation, the present symptom was qualified within the colour green indicating a low risk of serious illness. The NICE guideline itself includes a traffic light system chart for febrile illness in children under the age of five for identifying the likelihood of serious illness. The Schmitt-Thompson protocol does not use such a system. Therefore this protocol was also converted into a traffic light system chart by stratification of disease symptoms, allowing comparison of features.

Results

Search Results

Out of the 878 apps retrieved from initial searches in the Apple App Store and the Google Play Store, 6 applications matched our prespecified selection criteria. From these 6 apps, 3 apps were additionally excluded given the lack of paediatric fever focus. The results of these applications will therefore not be discussed within this section. The include apps are the Kinsa app, the FeverFriend app and the Feverapp (17, 18, 19).

Evaluation against the NICE guideline

All three applications were evaluated against the NICE guideline using stratification of disease symptoms into five categories, i.e. fever, skin condition and colour, respiration, circulation and hydration, general condition and activity. Within the NICE traffic light chart, 51 disease symptoms were identified. Each symptom was screened for within the apps' algorithm. Presence of the symptom and stratification within the accurate level of risk according to the guideline, was identified as a match between app and guideline. Absence of the symptom or stratification within an inaccurate level of risk was identified as a no match between app and guideline.

The FeverFriend app showed the highest adherence to the guideline with a total of 26 matches (51%), closely followed by the Kinsa app with a total of 24 matches (47%) and the FeverApp app with a total of 20 matches (39%), respectively. A detailed summary of the results can be found in Table 2.

Evaluation against the Schmitt-Thompson protocol

All three applications were equally evaluated against the Schmitt-Thompson protocol, using stratification of disease symptoms into the five categories. Within the Schmitt-Thompson traffic light chart, 42 disease symptoms were identified. The Kinsa app showed highest adherence to the protocol with a total of 28 matches (67%), followed by the FeverFriend app with a total of 21 matches (50%) and the FeverApp app with a total of 12 matches (29%), respectively. A detailed summary of the results can be found in Table 3.

MARS App Quality Scores

Looking at the first section and including all three applications, the average MARS overall quality score was 4.3 out of 5 with a standard deviation (SD) of 0.2 and a range of 4.0-4.4. Functionality was found to be the highest scoring domain (mean 4.5 [SD 0.4]; range 4.2-4.9), followed by aesthetics (mean 4.4 [SD 0.5]; range 3.9-4.8), engagement (mean 4.1 [SD 0.4]; range 3.7-4.4) and information (mean 4.0 [SD 0.4]; range 3.5-4.2) respectively. Subjective app quality, as a measure of rater satisfaction, was the lowest scoring MARS section (mean 3.5 [SD 0.8]; range 2.6-4.0). Likelihood for behavioural impact, being the third and final section, had an acceptable average score (mean 3.8 [SD 0.5]; range 3.3-4.1). With regard to the total quality rating within the

individual apps, the Kinsa app and the FeverApp app were equally the highest scoring mobile applications

(4.4 out of 5). Within the app specific subjective quality rating, we find similar scores when looking at the Kinsa app and the FeverFriend app (4.0 and 3.9 out of 5 respectively). The FeverApp however scores markedly lower in this section (2.6 out of 5). Independent ratings on the overall MARS total score and the overall subjective quality score of

Table 4: Summary of the Mobile App Rating Scale scores across the 3 reviewed apps.

MARS Scores and Intraclass correlation coefficient					
MARS domain	**Kinsa**	**FeverApp**	**FeverFriend**	**Mean [SD] 95% CI)**	**ICC (95% CI)**
Engagement	4.4	4.2	3.7	4.1 [0.4] (3.2 - 5.0)	0.90 (0.47 - 1.00)
Functionality	4.5	4.9	4.2	4.5 [0.4] (3.7 - 5.4)	0.84 (0.25 - 1.00)
Aesthetics	4.4	4.8	3.9	4.4 [0.5] (3.2 - 5.5)	0.89 (0.46 - 1.00)
Information	4.2	3.5	4.2	4.0 [0.4] (3.0 - 5.0)	0.95 (0.68 - 1.00)
Total MARS Quality	4.4	4.4	4.0	4.3 [0.2] (3.7 - 4.8)	0.82 (0.17 - 1.00)
Subjective Quality	4.0	2.6	3.9	3.5 [0.8] (1.6 - 5.4)	0.95 (0.71 - 1.00)
Likelihood of Behavioral Impact	4.1	3.3	4.1	3.8 [0.5] (2.7 - 5.0)	0.95 (0.74 - 1.00)

Abbreviations:
MARS = Mobile Application Rating Scale; SD = standard deviation; ICC = Intraclass correlation coefficient.

the three applications demonstrated an appropriate level of interrater reliability (ICC = 0.82, 95% CI 0.17-1.00 and ICC = 0.95, 95% CI 0.71-1.00 respectively) (17). A detailed summary of the quality assessment scores and ICC of the included apps is presented in Table 4.

Discussion

To our knowledge, our study is the first to evaluate the performance and accuracy of mobile self-assessment applications assisting parents in the assessment of fever in their children. Our research therefore provides insight into whether a mobile application with parental decision support system has the potential to reduce pressure on primary and secondary care clinics in a safe and patient-friendly way.

To date, several studies have investigated the usefulness of mobile applications in different fields of healthcare. Self-assessment tools, based on computerized clinical algorithms have been described in the literature for primary care for children and adults through the use of symptom checker applications (8). Semigran et al. evaluated the triage accuracy of 15 symptom checker tools. In 57% of cases an 'appropriate level of care' was found (20). However, the performance of the mobile application varied depending on clinical severity. Patient evaluations in emergency care showed a markedly higher percentage of correct triage advice (80%) than those for which self-care was used (33%) (20). Verzantvoort et al. assessed a self-triage app for several symptoms in the context of acute primary care. An accurate level of care was advised in 81% of cases (8). Our study identified remarkably lower percentages. On average the applications advised an appropriate level of care in 46% and 48% of cases, respectively, compared with the NICE guideline and the Schmitt-Thompson protocol. Self-assessment tools as described in the studies mentioned above, have been promoted as a means of reducing unnecessary office visits. However, analysis of these tools in the past has identified a risk-adverse tendency. On one hand over-triaging may improve patient safety since it lowers the chances of missing a red flag symptom. On the other hand overly cautious triaging by the mobile applications might encourage consultation and cause unnecessary care seeking. In our study, no tendency towards risk adverseness could be identified within the FeverFriend App or the FeverApp app. The Kinsa app however, did show an over-triaging propensity.

Fever management advice was compared to two sets of guidelines. The NICE guideline concerning fever management under the age of 5 on one hand and the Schmitt-Thompson protocol concerning fever under and above the age of 3 months on the other. The National Institute for Health and Care Excellence is an executive non- departmental public body in England known to publish updated, EBM recommendations guiding health promotion in a (cost-) effective and safe manner (15). Schmitt-Thompson Clinical Content is the leading source of telephone triage guidelines in North America, providing high quality protocols offering practical evidence-based decision support (16). In our study 3 mobile applications were assessed. The FeverFriend App was developed under direction of the University of Pecs in Hungary. The FeverApp was similarly developed under direction of the University of Witten in Germany, in cooperation with the Professional Association of Paediatricians (BVKJ) and the German Society for Paediatric and Adolescent Medicine (DGKJ). The Kinsa App was created by Kinsa Inc., a health technology company headquartered in the United States of America. Both European and American applications were thus included within this study. Identical selection of a European and American guideline limited the bias in results that arises from comparing an American app with a European guideline or vice versa.

Following the Cicchetti guidelines for interpretation of ICC interrater agreement measures, we identified an excellent level of IRR on all MARS domains (21). However, since all calculated limits of the 95% confidence intervals reach below 0.75, the ICC should not be regarded as truly excellent. Based on the ICC results for information, subjective app quality and likelihood of behavioural impact, we can conclude the IRR to be good to excellent. This is because the true ICC value supposes to land on any point within the ranges identified by the 95% CI. For the engagement and aesthetics domain we conclude the IRR to be fair to excellent. For the functionality domain and the total MARS quality score, we conclude the IRR to be appropriate (14).

There are several limitation of the current study. First, despite thorough search of earlier specified digital distribution platforms, we cannot be certain that all currently available mobile applications with a parental decision support system handling fever in children were identified and assessed within this study. Second, only one out of the 3 applications (the FeverFriend app) identified 3 levels of risk as shown in table 2 and 3. The Kinsa app identified 5 levels of risk and the FeverApp app generally acknowledged only 2 levels of risk ('do not call a doctor' or 'call a doctor'), with exception of few alarming symptoms ('call a doctor now'). In the process of implementation only the advices with clearly urgent implication, for example 'call a doctor now', were triaged red. This process of stratification might have played a role in the over- and under- triaging tendency identified within the Kinsa and FeverApp app respectively. Third, we did not measure the user's intention to follow the advice given by the app. Therefore, no insight in the efficiency of gatekeeping tools and the effect it has on healthcare seeking behaviour in real life was obtained.

Future app design in the field of medical decision making should ensure that established, evidence-based guidelines are incorporated upfront, that reliable and transparent algorithm design is used (e.g. providing the decision tree that is used as metadata), ensure regular updates and reviews in case guidelines change, while focusing on quality assurance and user friendly design in the interface of the app.

Conclusion

E-health is expected to get a more prominent place within healthcare. Given the low operational cost, self-triage- and management mobile applications could lower the pressure on the healthcare system by empowering parents to self-manage in a safe and cost- effective manner (8). However, there are some potential negative aspects such as changing the waiting room prevalence and pretest probability due to self-triage, data privacy concerns, inequity in access to these apps as well as legal and ethical concerns.

Moreover the limited availability of evidence based parental decision support applications for fever in young children remains a weakness. In this study we identified the Kinsa app as the most evidence based mobile application at present.

The authors have no conflicts of interest to declare with regard to the topic discussed in this manuscript.

References

1. Nelson Textbook of Pediatrics [Internet]. [cited 2024 Sep 8]. Available from: https://elsevier-elibrary.com/contents/fullcontent/15188720/epubcontent_v2/OEBPS/B978143770755700169X.htm
2. Villarejo-Rodríguez MG, Rodríguez-Martín B. Parents' and primary caregivers' conceptualizations of fever in children: A systematic review of qualitative studies. Nurs Health Sci. 2020 Jun;22(2):162–70.
3. Thompson AP, Nesari M, Hartling L, Scott SD. Parents' experiences and information needs related to childhood fever: A systematic review. Patient Educ Couns. 2020 Apr;103(4):750–63.
4. Walsh A, Edwards H. Management of childhood fever by parents: literature review. J Adv Nurs. 2006 Apr;54(2):217–27.
5. Chefdeville E, Pages AS. Parental management of children's fever: Assessment of knowledge and use of health record information. Arch Pediatr. 2019 Jul;26(5):275–81.
6. Green R, Jeena P, Kotze S, Lewis H, Webb D, Wells M. Management of acute fever in children: guideline for community healthcare providers and pharmacists. S Afr Med J. 2013 Sep;103(12):948–54.
7. Neill S, Roland D, Jones CHD, Thompson M, Lakhanpaul M. Information resources to aid parental decision-making on when to seek medical care for their acutely sick child: a narrative systematic review. BMJ Open. 2015 Dec;5(12):e008280.
8. Verzantvoort NCM, Teunis T, Verheij TJM, van der Velden AW. Self-triage for acute primary care via a smartphone application: Practical, safe and efficient? PLoS One. 2018;13(6):e0199284.

9. Buechi R, Faes L, Bachmann LM, Thiel MA, Bodmer NS, Schmid MK, et al. Evidence assessing the diagnostic performance of medical smartphone apps: a systematic review and exploratory meta-analysis. BMJ Open. 2017 Dec;7(12):e018280.
10. Jazayeri SMHM, Jamshidnezhad A. Top Mobile Applications in Pediatrics and Children's Health: Assessment and Intelligent Analysis Tools for a Systematic Investigation. Malays J Med Sci. 2019 Jan;26(1):5–14.
11. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth. 2015 Mar;3(1):e27.
12. Schmeelk S, Davis A, Li Q, Shippey C, Utah M, Myers A, et al. Monitoring Symptoms of COVID-19: Review of Mobile Apps. JMIR mHealth uHealth. 2022 Jun;10(6):e36065.
13. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A Systematic Review of Smartphone Applications Available for Corona Virus Disease 2019 (COVID19) and the Assessment of their Quality Using the Mobile Application Rating Scale (MARS). J Med Syst. 2020 Aug;44(9):164.
14. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155–63.
15. Overview | Fever in under 5s: assessment and initial management | Guidance | NICE [Internet]. [cited 2024 Sep 8]. Available from: https://www.nice.org.uk/guidance/NG143
16. Office-Hours Telehealth Triage Protocols User's Guide Office-Hours Telephone Triage Protocols User's Guide 2023 Office-Hours Telehealth Triage Protocols User's Guide. 2009 [cited 2024 Sep 8]; Available from: www.cleartriage.com.
17. Feel Better Faster with the Kinsa App | Kinsa Health [Internet]. [cited 2024 Sep 8]. Available from: https://home.kinsahealth.com/kinsa-app
18. FeverFriendTM | Welcome [Internet]. [cited 2024 Sep 8]. Available from: https://feverfriend.eu/
19. Fever App [Internet]. [cited 2024 Sep 8]. Available from: https://www.feverapp.de/en/
20. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ. 2015 Jul;351:h3480.
21. Cicchetti D V. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. Psychol Assess. 1994;6(4):284–90.