Baby CROINC: An Online, Crowd-Based, Expert-Curated System for Monitoring Child Development

Ayelet Ben-Sasson
University of Haifa, Israel
asasson@univ.haifa.ac.il

Eli Ben-Sasson
Technion, Haifa, Israel
eli@cs.technion.ac.il

Kayla Jacobs
Technion, Haifa, Israel
kayla@cs.technion.ac.il

Eden Saig
Technion, Haifa, Israel
edens@cs.technion.ac.il

ABSTRACT

Baby CROINC (CROwd INtelligence Curation) is an online early-childhood development tracker designed to be both personalized and objective. To meet these goals, we rely on Curated Crowd Intelligence (CCI), a process in which experts curate personalized inputs to connect with the crowd's aggregate data, providing parents with objective and personalized feedback on their children's development. In this paper, we describe Baby CROINC's design, with a focus on CCI, and assess the extent to which it meets its design goals of objectivity and personalization.

In Baby CROINC, parents create a diary by adding developmental milestones to a timeline. Visual statistics are presented per milestone. Expert curators clarify, merge, and classify milestones which are new to the system.

Diary personalization was evident through users' rich and diverse milestone choices, and by the continuous system increase in new canonical developmental concepts. Findings demonstrate the objectivity of the crowd-based percentiles extracted from Baby CROINC, based on consistency of developmental differences in preterm vs. fullterm and boys vs. girls with established research, and the correlation between medians reported in our system and those appearing on the U.S. Centers for Disease Control and Prevention's Milestones webpage.[1]

CCI led to a dramatic increase in users' ability to view crowd-based statistics, indicating that CCI is critical for enabling objectivity while maintaining personalization.

Author Keywords

early child development; crowd sourcing; health informatics

[1] www.cdc.gov/ncbddd/actearly/milestones, adapted from [27, 12].

PervasiveHealth '17, May 23–26, 2017, Barcelona, Spain

ACM ISBN 978-1-4503-6363-1/17/05…$15.00
https://doi.org/10.1145/3154862.3154887

INTRODUCTION

Developmental screening of young children is important for detecting developmental delays. This can facilitate referral for early intervention [23] when support resources are more effective [6], as well as reassurance and education of parents of normally developing children. Early childhood developmental screening is typically performed by pediatricians and baby-wellness specialists, and requires the administration of complex, expensive, and time-consuming standardized tests; properly interpreting the results requires additional time and expertise [19]. Without expert help, the average parent faces challenges both in administering tests and in deciphering their outcomes. These challenges are particularly relevant to developing countries and lower socio-economic strata, where they become more acute due to the limited availability of screening and patient-provider time [7].

New technologies provide new opportunities for parents to monitor and learn about their children's development. Many parents of young children search the Internet for information regarding their children's health and development, and use social media to record developmental milestones [17, 25, 30, 4]. However, it is difficult for parents to obtain information that is both *objective* and based on quantitative data while still being *personalized* and catering to specific, parent-driven interests and concerns. With a few notable exceptions described below, there is a shortage of tools that lie between the two extreme points of standardized tests, which are objective but not personalized, and parenting forums and social media, which are personalized but not objective.

Baby CROINC and CCI

Baby CROINC (CROwd INtelligence Curation) is a free early childhood development tracker available as a website[2] and as iPhone[3] and Android[4] mobile apps. It aims to resolve the tension between *objective* feedback and *personalization*, using what we call Curated Crowd Intelligence (CCI).

[2] baby.croinc.org

[3] itunes.apple.com/lt/app/baby-croinc/id1185899162

[4] play.google.com/store/apps/details?id=com.babycroic.croinc

In Baby CROINC, the parent or caregiver creates a custom developmental diary for their child via a sequence of *personalized*, age-dated milestones like “first smile” at five weeks old or “eating with a spoon” at seven months old. For each diary-recorded event the system provides the user with *objective* statistical information about the milestone associated with it. In addition, parents/caregivers can view ages predicted for achievement of future milestones.

This parent-driven system achieves the goal of *personalization* because caregivers choose what to record and track, and are encouraged to author new milestones as well as select from suggestions of common existing milestones. Simultaneously, information displayed regarding each milestone is quantitatively based on aggregated statistics gathered from all users’ children, thereby achieving the goal of *objectivity*.

To seamlessly integrate these two seemingly contradictory goals, and to ensure the statistics’ high quality, the CCI approach involves child-development experts cleaning and *curating* the personalized parent-authored data, as explained later in the “Design Goals and Curated Crowd Intelligence (CCI)” section.

Related Work

When designing Baby CROINC, we have benefited from lessons learned from existing child development educational resources and prior interactive and web-based developmental trackers, several of which are described below.

Static, curated data Static, curated information on child development is already available online, and frequently sought out by parents of young children [17, 30, 15]. For example, the United States Centers for Disease Control and Prevention (CDC) publishes lists of milestones for children aged 0-5 years, categorized by developmental domains, in an effort to increase parental knowledge of age-appropriate milestones and red flags. These lists are based on two developmental guideline books for caregivers by the American Academy of Pediatrics [27, 12].

Parenting forums Many parents are active on various online parenting forums, social media networks, and Q&A communities, asking questions and raising concerns of their own. Sometimes health care professionals and developmental experts respond as well. But overall, this online interaction is limited, biased, and narrow in terms of the data provided by parents and experts [3, 22]. Our CCI approach is further motivated by existing work that stresses the need to improve the quality of general crowd-sourced wisdom using real-time expert guidance [5, 21].

Popular apps and websites Many technologies allow parents of young children to track their infant’s activities, schedule, and/or physical growth. Examples include *Trixie Tracker*™ [20], *Text4Baby*™ [14] and *MyPreemie* [8]. Baby CROINC differs from these by (i) expanding its focus beyond the physical aspects of early childhood development; and (ii) its unique CCI approach.

An interesting use of hand-held electronic devices to obtain precise and objective data regarding early childhood development is reported in [9]. In that experiment, a cohort of 40 expectant mothers was trained to identify important motor milestones and report them on a Palm Pilot. This provided excellent quantitative results but, in contrast with Baby CROINC, required rigorous training and does not support user-driven authoring of new milestones.

@BabySteps is another influential system, and the first interactive Twitter-based developmental tracker [29]. This system prompts parents via Twitter messages to answer age-relevant developmental questions drawn from the Ages & Stages Questionnaire [28] screening tool. It yields quantitative results [29, 18] but is not as personalized as Baby CROINC.

PatientsLikeMe® (PLM) is a popular platform sharing conceptual and technological similarities with Baby CROINC, though it does not focus on early childhood development [11]. PLM is a social network for people with common health conditions, where users share their symptoms and treatment schedules, mainly through built-in surveys, and track their health status relative to the crowd. Numerous visualization tools assist users in interpreting the data. Research indicates that PLM is associated with increased patient health knowledge and improved well-being and decision-making capabilities [31]. Baby CROINC addresses a similar need but caters to parents of young children, offering developmental knowledge and tracking that can similarly empower them and inform their decisions.

Paper Organization

In the next section we explain our two design goals and CCI in more detail, followed by a fuller description of Baby CROINC’s features and usage patterns. In the Assessment section, we examine the extent to which our goals are achieved and the impact of CCI. We conclude with a Discussion of our findings and thoughts for future work.

BABY CROINC DESIGN GOALS

The main purpose of Baby CROINC is to help parents and caregivers track and understand their child’s unique early developmental trajectory in a *personalized* and *objectively* meaningful way, achieving both these seemingly-divergent goals through CCI.

Design Goal: Personalization

Standardized tests achieve high levels of validity, reliability, and replicability. These advantages come at the price of (i) specificity to particular child development histories and trajectories, (ii) rigidity of the test format regardless of changing times and ethnic/cultural environments, and (iii) considerable research effort required to calibrate test scores and interpret them [26, 16]. Personalized, non-standardized assessments are more flexible and can adapt quickly to changing environments and parent interests. (See, e.g., [1] and references within for more discussion of the ongoing “standardized vs. personalized” testing/assessment debate.)

Our Baby CROINC system offers a high level of personalization by encouraging users to choose and author the milestones recorded for their children (see Figure 1). The diaries thus

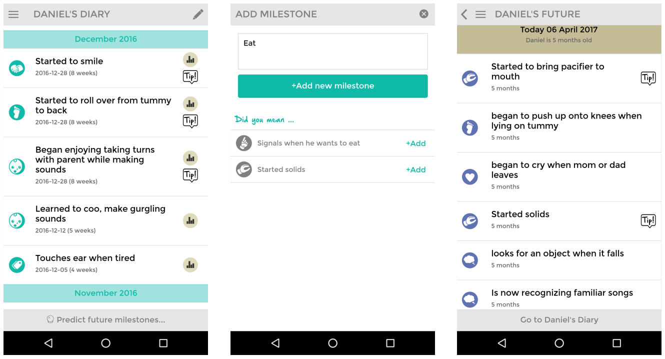

Figure 1. Screenshots from a developmental diary in the Baby CROINC mobile app. On the left is the personal diary display; in the middle is the interface for adding a new milestone to a diary, including autocomplete suggestions towards the bottom half of the screen; on the right are a list of future milestones the child can likely anticipate.

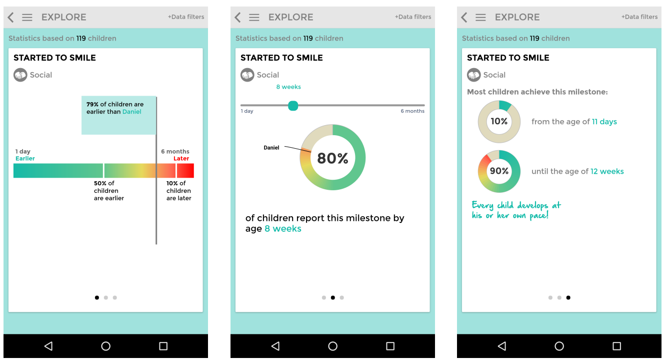

Figure 2. Exploring statistics for the "started to smile" milestone with the Baby CROINC mobile app. In the (simulated) example above, the child is late for this milestone (21st percentile, with sample size of $n =$ 119), having a reported age of 2 months. Notice the implementation of the system's design goals (see section "Baby CROINC Design Goals"): a *personalized* graphical display, providing *objective*, non-parametrized, percentile-based statistics.

cater to parents' and caregivers' specific interests and concerns. Additionally, the statistical data is presented to users in a child-focused manner (see Figure 2), so that the parent can easily interpret what the data says about his or her child.

Design Goal: Objectivity

The level of personalization we offer in Baby CROINC could lead to a biased and subjective assessment of child development. To compensate for this risk, Baby CROINC is also based on a quantitative research approach, pooling together data from across children's individual diaries, and presents only non-parametric statistics to avoid making any assumptions about the underlying distribution.

Design Goals and Curated Crowd Intelligence (CCI)

Personalization and objectivity lead to seemingly contradictory data collecting procedures; personalization means that users log whatever milestones they find interesting or relevant, while objectivity and quantitative analysis typically require data from large populations that answer the same set of questions.

We have resolved this tension in Baby CROINC with a new process in which experts curate the data provided by users (the crowd). Caregivers may add to their child's diary any milestone written in whichever words or style they like, in keeping with the goal of personalization. In order to provide objective statistical feedback, expert *curators* work behind the scenes to connect the milestone to others that are conceptually synonymous, even if the text is different. For example, different children might have milestones in their diaries with unique texts like "began walking," "he started walking," "Sophie is walking already!"—all of which collectively draw from a shared pool of statistical data. We call this collection of conceptually synonymous milestones with pooled statistics *canonical milestone concepts*.

DESCRIPTION OF Baby CROINC

User Input and Engagement

Baby CROINC aims to provide users with an intuitive and engaging interface. Parents and caregivers simply add developmental milestones their child currently or retrospectively achieved to a dated timeline. Milestone texts are short phrases with a mean length of 6.90 words (median = 6.00 words, SD = 3.56 words, range = 1–27 words), and can be either selected from a list of suggestions or authored uniquely by the user. In addition, users may view anticipated upcoming future milestones as reported by older children, as well as milestone-focused parenting tips provided by other caregivers. See Figure 1 for screenshots of the interface.

Automatic e-mail notifications update users after extended periods of inactivity, encouraging them to update their children's diaries with a suggested list of age-relevant milestones. Additionally, users are notified when statistics are available for milestones previously without viewable statistics.

System Output

For each milestone, such as "started crawling," Baby CROINC calculates aggregated information from all other users who added that same milestone to their children's diaries. Simple percentile-based statistics are presented to users using an intuitive graphical user interface, including the child's percentile and the overall age distribution for all children (see Figure 2). The statistics are percentile/median-based as opposed to mean-focused, to make them less sensitive to undue influence from statistical outliers (severely delayed or advanced children whose ages lie at extreme values)

Statistics are only shown for milestones which have appeared in at least five children's diaries. For milestones below this threshold, an e-mail notification is sent to users once the threshold is reached, alerting them to the availability. The chosen cut-off balances (i) the desire to provide statistics for as many milestones as possible without an overly long delay, against (ii) the need for the statistics to be numerically meaningful and based on a non-trivial sample size. We always provide users with the sample size so they can evaluate the strength of the statistics, to distinguish between milestones with few vs. many children. As Baby CROINC continues to expand, the sample sizes for most milestones will increase as they are added for more children, and statistical relevance will increase.

System Initialization

When the system was launched, before any users joined, 252 canonical milestone concepts were seeded for the introductory Baby CROINC suggestion list. These milestones were drawn from a list of published developmental milestones from the U.S. CDC (described above in the Related Work section).

Curated Crowd Intelligence (CCI) Expert Review Process

The high quality statistical output provided to users is dependent on behind-the-scenes expert curation. The team of experts performing this regular curation has a background in child development and were trained to consistently and reliably curate milestone texts. The first author, a research academic specializing in child development, conducts periodic reviews of milestone classification to ensure quality and to provide team members with additional feedback.

When a user adds a milestone to their child's diary via a system suggestion of an existing milestone, that child's age is automatically included in the milestone's pool of statistics. However, for new milestone texts authored by users, expert curation is required.

If the milestone's underlying idea already exists in the system as a canonical milestone concept, but the user's milestone uses text that varies somewhat in language, the expert *merges* the milestone with the existing canonical milestone concept to associate them together. For example, a new milestone text "Bobby can ride a three-wheeled bike" could be merged with the existing "learned to use a tricycle." This reduces redundant milestone concepts and provides aggregated statistics, independent of minor variations in text language.

For milestones that represent an entirely new concept in the system, the expert approves the creation of a new canonical milestone concept. Before it is added to the system-wide list of milestone suggestions, the expert may lightly edit the text if needed to fix typos, remove any references to gender or the child's name to make it more universal, translate to English if originally written in a non-English language, etc. These language edits always stay true to the original intended meaning of the milestone text, but make it more accessible to other users who may want to add it to their children's diaries. Additionally, the new canonical milestone concept is assigned one or more developmental domain categories like gross motor, fine motor, language comprehension, etc. (described further in the Developmental Domains of Milestones section below).

If the milestone does not qualify as a developmental milestone (e.g. "first visit to Grandma"), or is gibberish, the expert flags the milestone to ensure it is not suggested to other users or included in overall developmental statistics.

This expert curation process ensures clean, reliable, aggregated statistics for milestones and high-quality suggestions, while preserving the user's unique, personalized language choices and authorship within their own diaries.

Current User Base

The current user base (as of December 27, 2016) comprises 490 users who completed the registration process for a total of 518 children. Users were recruited through advertising on social media, especially targeting groups supporting parents of young children, and through media exposure about the project. A small number of users were recruited as participants in studies performed by university students and researchers in child development, affiliated with the first author. Most (95.10%) users registered one child, and almost all (99.03%) identified themselves as the children's biological mothers.

The average age of children at registration was 16.15 months (median = 10.37 months, SD = 18.34 months, range = 3 days – 13.29 years; see Figure 3). As fitting for the early childhood focus of the project, 95.75% were under age 4 years, 91.70% were under age 3 years, and 81.27% were under age 2 years. We found the few (4.25%) older children in the system tended to fall into one of two categories: (1) older children who were the siblings of young children, added by users who focused primarily on the younger sibling(s) but also created diaries for the older sibling(s); and (2) children with significant developmental delays, whose parents continued to closely monitor their milestones past the young ages at which typically-developing children are routinely monitored. It is important to note that since the statistics shown to users are based on percentiles, which are insensitive to outliers, any developmentally-delayed children do not noticeably skew statistics for typically-developing children.

About half (50.97%) of children were boys, and the average pregnancy length was 38.90 weeks (median = 39 weeks, SD = 1.94), with 8.16% of the sample born preterm (defined as <37 weeks of gestation). For those who provided extra elective information, on average birth weight was 3.25 kg (median = 3.27 kg, SD = 0.63 kg, range = 1.30–5.19 kg), 3.28% of children were part of a multiple-order pregnancy like twins or triplets. A majority (67.67%) were the first-born children in their families, and another 19.23% second-born.

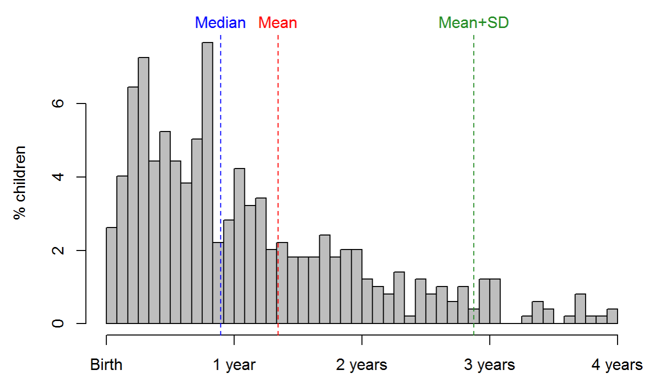

Figure 3. Child ages at registration (0-4 years).

User Activity Patterns

Counted together, users added a total of 5,894 milestones to their children's diaries, corresponding to 537 canonical milestone concepts (for example, "started walking" and "began to walk" counted as two milestones but as a single canonical milestone concept). The mean number of diary milestones per child, after accounting for some deletions, is 10.99 (median = 5.00, SD = 23.63, range = 1–142, see Figure 4).

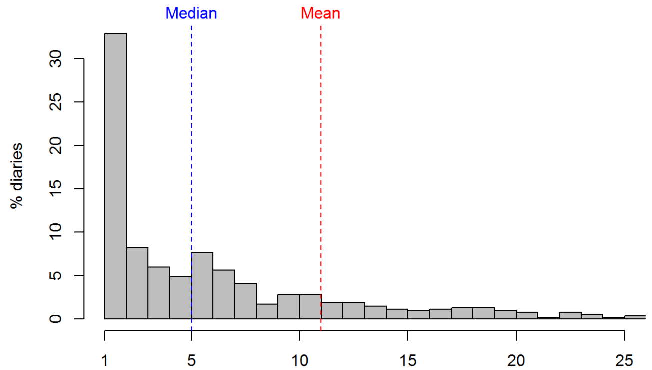

Figure 4. Number of milestones per child's developmental diary. An additional 8.97% of diaries have over 25 milestones, up to 142 (not shown).

Users added milestones to their diaries in any of three main ways (see Figure 5):

1. selecting from a list of suggested existing milestones (58.16% of milestones);
2. entering their own unique, original milestone text (16.74%); or
3. starting to type their own milestone text and then selecting an existing milestone from an "auto-complete" suggestion matching their partial text to a list of existing milestones (14.93%).

No matter the method, users always recorded the age (or, equivalently, date) at which their child achieved that milestone.

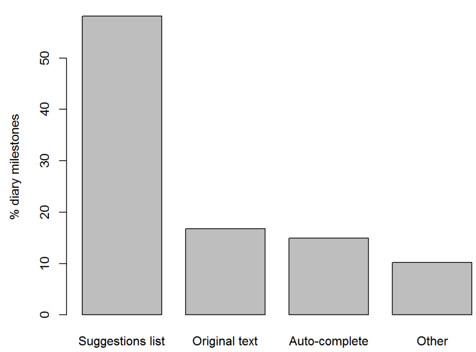

Figure 5. The various methods users chose to add milestones to their children's diaries. Users mostly add milestones from the accumulating suggestion lists.

The age at which a milestone was achieved was typically not the same day the milestone was recorded in Baby CROINC—in fact, only 4.13% of milestones were recorded the same day, though 10.26% were recorded within a week. The lag time between a milestone's achievement and its addition to the diary could be because the user reported a child's milestones retrospectively for ages prior to Baby CROINC registration, or because they only thought to record it later (perhaps after seeing it on the list of suggested milestones). The median time lag is 2.47 months (mean = 5.96 months, SD = 11.10 months).

As is common for websites and mobile apps, most users did not subsequently return to Baby CROINC to add additional milestones to their children's diaries after their initial visits. (Users may have returned to view the existing diary and statistics while not entering additional milestones, but such behavior is not currently logged, and hence is not included in the visit count.) However, 22.10% of users *did* return for additional diary entries; of these, over half came back for a single additional visit, and the rest up to 43 times. The overall user activity time range—that is, the time between the dates of the user's first and last/latest active visit—averaged 73.56 days (median = 39 days, SD = 88.19 days, range = 1 day – 1.25 years; see Figure 6) with 88.70% within 6 months.

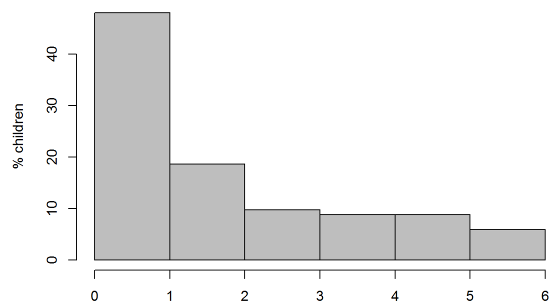

Figure 6. User activity time range (0–6 months).

ASSESSMENT

We now examine how Baby CROINC performs with regard to the two design goals of personalization and objectivity, and explore the impact of CCI.

Assessment of Personalization Goal

Users personalized their children's diaries in several ways.

Milestone Choices

Which milestones to include in their children's diaries was entirely up to the user, and their choices are highly diverse. For example, consider the three most popular milestones in the system:

1. "began to smile at people" in 115 (22.20%) diaries,
2. "started walking alone" in 100 (19.31%) diaries, and
3. "started crawling" in 96 (18.53%) diaries.

Of the 518 children in the system, only 9 (1.74%) had all three milestones in their diaries, while 326 (62.93%) did not have *any* of them. 15 (2.89%) children had (1) and (2) but not (3); another 15 (2.89%) children had both (1) and (3) but not (2); and 31 (5.98%) children had both (2) and (3) but not (1).

Developmental Domains of Milestones

Part of the user's choices about which milestone to record is also which aspect of child development to focus on, whether more physical or communicative. Figure 7 shows the distribution of all diary milestones as categorized by developmental domain. Gross motor (e.g., "started crawling") is the most popular domain, followed by fine motor ("started eating with a spoon"), cognitive ("follows things with eyes"), speech ("began to babble"), social ("shows affection to parents"), emotional ("shows fear of strangers"), and language comprehension ("follows simple instructions").

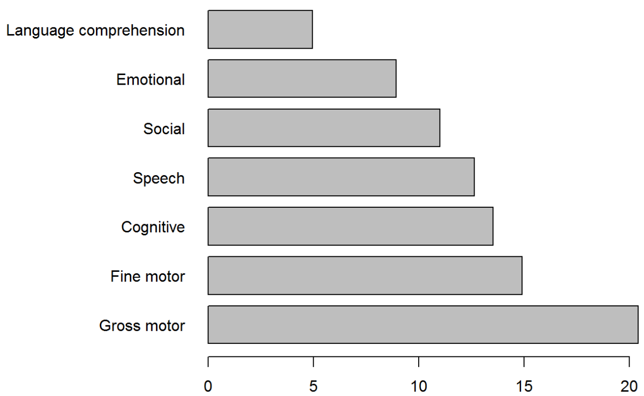

Figure 7. Percentages of diary milestones by developmental domain. About a third of diary milestones are devoted to motor development.

Additional lower-frequency developmental domains, not shown in the figure, include self-care activities ("began to help undress himself"), non-verbal communication ("points at what she wants"), events ("first day of nursery school"), regulation ("sleeping through the night"), medical ("flu vaccination"), and behavior problems ("started throwing food intentionally"). A few parents also chose to record attributes ("loves music") and physical features (such as weight measurements).

Ages of Milestones

Another part of the user's choices about which milestone to record is also which age to focus on. Most milestones were reported during the child's first year of life: the overall median milestone age is 229 days (SD = 308.02, range = birth–10.4 years, see Figure 8), and 97.17% of milestones were reported for under the age of four years. As expected, this milestone age distribution followed the general shape of the distribution of ages at registration (Figure 3).

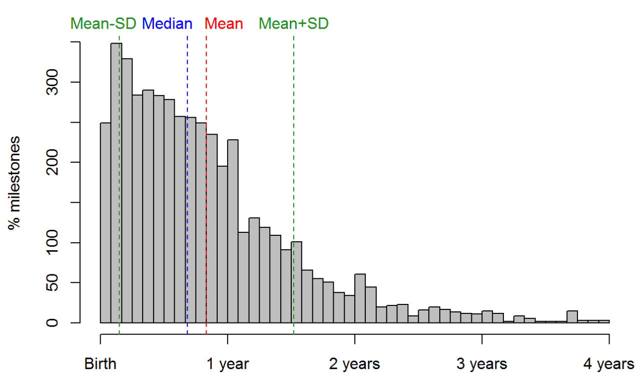

Figure 8. The child ages (0–4 years) for which milestones were reported.

For those children with at least two diary milestones, the average age range of reported milestones—i.e., the time gap between the youngest to the oldest milestone within the diary—is 9.52 months (median = 6.08 months, SD = 12.17 months, range = 1 day – 10.42 years; see Figure 9). Of course, the overall age of the child[5] is important to consider when understanding this diary age range. For example, a 6 month diary range for a 6-month-old child covers 100% of the child's lifetime, whereas the same 6 month diary range for a 2-year-old child covers only 25% of the child's lifetime. On average, the diary covered 56.32% of the child's lifetime (median = 58.50%, SD = 27.66%, range = 1.08 – 100%; see Figure 10).

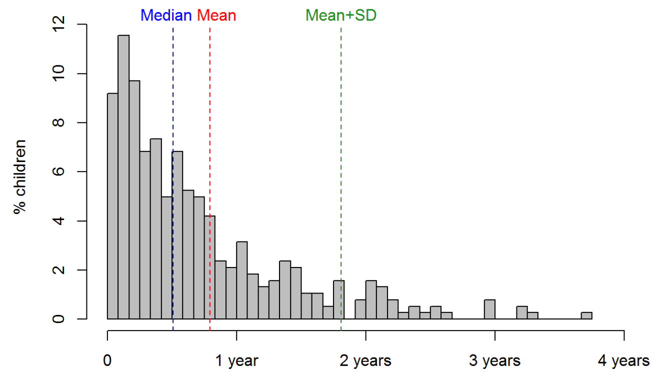

Figure 9. The range of time covered within a child's diary.

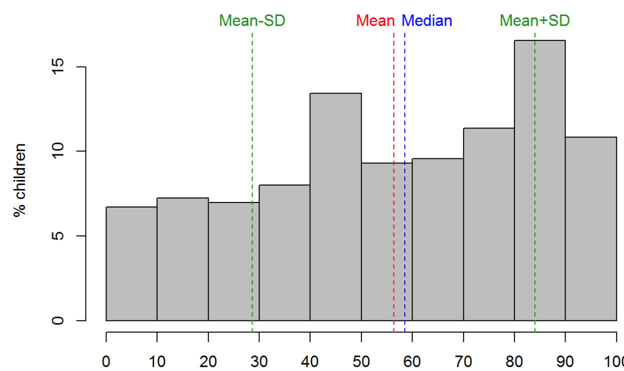

Figure 10. The percentage of a child's life covered by his or her diary.

Assessment of Objectivity Goal

Internal Validity for Baby CROINC Percentiles

Child development research (detailed in the Discussion section) shows developmental differences between preterm children and full-term children, as well as between boys and girls. We analyzed Baby CROINC for these patterns. An independent samples t-test showed that preterm children (defined as being born before 37 weeks of gestation) had significantly lower median percentiles, relative to full-term children, in the gross motor (22% difference, $t(305) = 4.10$, $p \leq 0.001$) and social (17% difference, $t(238) = 2.70$, $p = 0.008$) domains. Independent samples t-tests for the developmental domains, indicated that girls had higher median percentiles than boys in the social domain (10% difference, $t(238) = 2.76$, $p = .006$).

[5]We calculated a child's overall age at the date of the last diary activity, thus preventing skewing from the children's natural continuation of aging even after the parent has ceased or paused using the system to record milestones. Note the date of the last diary activity is always later than (or occasionally, the same as) the date of the last recorded milestone, as milestones are always recorded as having happened in the past (or occasionally, the present day).

External Validity for Baby CROINC Percentiles

For the 252 canonical milestone concepts in Baby CROINC which corresponded to published CDC milestones (described previously in the System Initialization section), we investigated the average age difference between the published CDC ages and the children's ages in CROINC for that milestone.

We expected a skew towards positive age differences—meaning the Baby CROINC milestone was achieved earlier (at a younger age) than the published CDC age. This is because the CDC age does not indicate the age *at* which the milestone was achieved (as the Baby CROINC age does) but rather the age *by* which the CDC recommends pediatricians start being concerned, if the milestone was not achieved anytime earlier. Additionally, CDC ages are purposefully published to correspond to typical well-baby visit ages (2, 4, 6, 9, 12, 18 months and 2, 3, 4, 5 years), which may not correspond to the milestone's "natural" age range. On the other hand, any Baby CROINC children with developmental delays contribute towards a skew in the opposite direction.

The results show that the median difference between the CDC age and the Baby CROINC age, across all milestones in all developmental domains, is 2.96 months (SD = 8.40 months). See Figure 11 for age differences between CDC and Baby CROINC per domain. Spearman correlations were high (r = 0.72–0.95 across domains) and significant ($p < 0.001$) between the median age reported for milestones in Baby CROINC across children and the age cutoff set by the CDC for corresponding milestones in each of the seven main developmental domains.

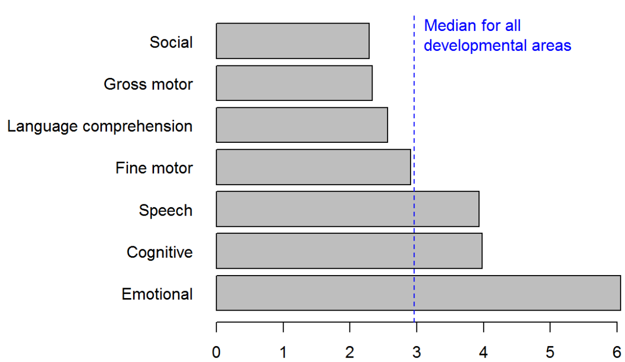

Figure 11. Median age difference between Baby CROINC age and CDC age, per developmental domain. The age difference is between 2.29 and 6.05 months with more complex domains showing the largest gap

Assessment of Curated Crowd Intelligence (CCI)

Learning New Concepts Using the Crowd

In exploratory research, the amount of meaningful collected data grows over time. Milestones reported by users in Baby CROINC are considered novel when they represent new canonical milestone concepts which didn't previously exist in the system, as opposed to simply a different language formulation of an existing context. For example, consider when a user adds a milestone like "David began to smile" when the "started smiling" canonical milestone concept already exists in the system—this does not add novel information to the system (even though the exact language of the milestone is new). However, novel information *is* added when a user creates a "David began interacting with videoconferencing" if no prior "began interacting with videoconferencing"-like canonical milestone concept existed. After some minor language edits from the expert curator, there is now a new canonical milestone concept in the system, available for other users to add to their own children's diaries.

As described in the User Activity Patterns section, the system was initialized with 252 canonical milestone concepts drawn from a published CDC list. Over time, user input increased the number of canonical milestone concepts in the system by 285, to reach the current total of 537—i.e., 238% of the original number of canonical milestone concepts. Figure 12 shows that the amount of meaningful knowledge in the system grows steadily as the users interact with it, demonstrating that our system is able to harness crowd intelligence to extensively create new, meaningful insights.

Of especial interest, the average slope of Figure 12 does not significantly decrease or plateau over time. This indicates that the system has not yet reached a "saturation" point. As a result, we anticipate that there is still a significant amount of childhood development milestone concepts that Baby CROINC has not yet captured. This evidence shows the importance of personalized, parent-authored and expert-curated recording systems like Baby CROINC, which greatly expand upon the existing knowledge base devised only by child development experts.

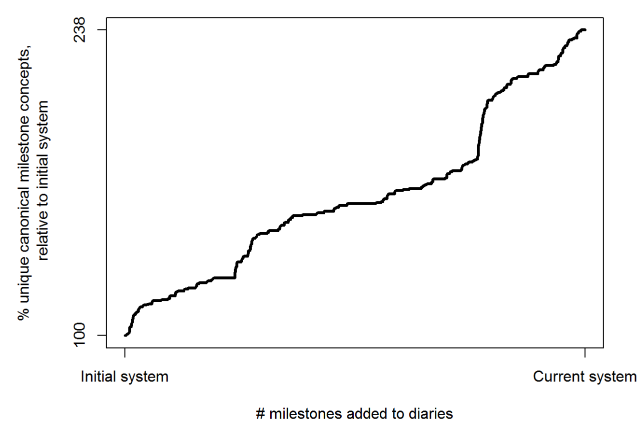

Figure 12. Exploratory data accumulation over time. The figure shows the growth of canonical milestone concepts in the system over time as users interact with the system, relative to the amount the system had when initially launched. The growth does not appear to have plateaued to a maximum value, indicating even more concepts are likely to be added in the future.

Making Milestone Statistics More Significant

Baby CROINC only shows its users statistics for milestones which have been recorded in the diaries of at least five children; below this threshold we consider the statistics as essentially meaningless. Merging equivalent milestones together makes it possible to display significant statistics, even when the data for each individual milestone text is relatively sparse.

For example, if essentially identical milestones like "started walking" and "began to walk" were treated as separate canonical milestone concepts, instead of being merged through the CCI process, then only 12.80% of milestones would contain enough data to provide users with significant statistics. By grouping equivalent milestone texts as canonical milestone concepts, the number of milestones which have significant statistics increased to 68.35%—an increase of 534%. This emphasizes the value of the CCI process.

Quantifying the Amount of Expert Curation Effort
Although very beneficial to Baby CROINC's ultimate quality, the CCI approach does add to the cost of running the system. To date, 1,013 milestone texts were created by the experts as clarifications for new user milestone texts (46% of all unique milestone texts in the system). In addition, 745 user-created milestones have been identified by the experts as equivalent to existing canonical milestone concepts. This level of curation requires regular review time by the trained expert team.

Analysis of system logs shows that curator work amounts to 1–2 hours during normal weeks, and 6–8 hours during peak weeks (e.g., time spent translating content when native speakers of a new language are introduced to the system). During normal weeks, the amount of curation work has a linear relation to the amount of user-created milestones in the system ($R^2 = 0.977$). Assuming that user-created milestones are the main factor driving the need for expert curation, linear regression analysis shows that curators spend an average time of 4 minutes and 11 seconds (± 4 seconds) on each user-created milestone.

DISCUSSION
Baby CROINC is a crowd-based platform for child development tracking and screening, which integrates expert-curated crowd data using a CCI process to achieve its two main design goals, *personalization* and *objectivity*. Baby CROINC facilitated the creation of personalized developmental diaries, as demonstrated by the rich array of milestone choices and developmental domains focuses seen in different users' diaries.

Parents' activity patterns show the value of allowing free selection and narration of a personal developmental profile that matches the stage, context and interests of each child and family. Parents tended to describe milestones in the context of specific activities, objects, or family (e.g., "Jim started playing with shape sorter"). From a health care perspective, this developmental documentation can highlight areas of particular concern for a given family and milestones of which parents may be unaware. At the same time, personalization makes it difficult to meaningfully compare diaries containing such different behaviors, ages, time frames, and language choices. Baby CROINC relies on the power of crowd wisdom in overcoming these challenges, as well as the CCI approach for identifying common milestones across diaries despite variations in parent language.

The most common milestones were in the gross- and fine-motor domains. This is probably because these domains are easily observable and constantly changing during the first year of life. Tracking motor milestones is important, as they are often the first warning signs observed at an early age even if the disorder is not neuro-muscular (e.g., [24]). However, it is important to increase parental awareness of milestones attained and expected in social-communication domains as well. This awareness is critical for identifying delays and helping parents meet and enrich their child's social-communication development. It is possible to facilitate the creation of developmentally diverse diaries by tweaking the pool of suggested milestones presented to parents; however, this may compromise the crowd-based approach Baby CROINC takes. The development of a design balancing both considerations is needed. These findings call for providing parents with more information regarding early childhood development, to promote awareness of age-appropriate development in the social-communication domain.

Of course, Baby CROINC is meant to support and augment, not replace, real-world health care professional evaluations when they are available. In the absence of such evaluations, whether because of time gaps between doctor check-ups or because of insufficient resources in developing or lower socio-economic communities, inaccurate information risks leading to unnecessary parental anxiety, and possibly resulting in false positive diagnoses of developmental delays—or worse, false negatives (where the child needs intervention but is not properly diagnosed). Considering these risks, research validating crowd-based percentiles is important to promote confidence in their accuracy. The findings presented in this paper start this process by demonstrating "real-world" patterns. The data in the Internal Validity for Baby CROINC Percentiles section show girls were in higher communication percentiles than boys, findings which are in line with traditional child development data [2]. Similarly, our findings about the lower developmental percentiles of preterm versus full-term children are in line with existing evidence [13]. In the External Validity for Baby CROINC Percentiles section, the Baby CROINC percentiles were also validated against the published CDC age norms for comparable milestones. While most Baby CROINC percentiles tend to present a more advanced developmental picture than the CDC, this is expected because the former records "first achieved" while the latter lists "when to worry" which is generally later. This difference was greatest for milestones in the emotional, speech, and cognitive domains. We offer two plausible hypotheses: (1) milestones in these domains have a higher threshold for clinical concern and a wider range of what is considered typical; or (2) parents may be less equipped to precisely characterize these abstract domains. Similar differences were demonstrated in previous research, where motor milestones were recorded earlier by mothers via electronic diary methods than they were observed clinically [9]. This pattern of differences between modes of assessment can be attributed to standardized assessment's goal of minimizing "false positives."

One of the potential threats to the objectivity of such a platform is its sample bias. However, our user base shows encouraging indications of sample representation, as seen in the equal distribution of boys and girls and the reported rate of preterm births amongst Baby CROINC children (8.16%), which corresponds to the population rate [10]. The correlations between

Baby CROINC percentiles and CDC-published norms in corresponding developmental domains indicate that the sample is not biased in terms of a greater percentage of children with a developmental delay. Further work is needed to look at sample representation across other demographic features in a diverse population.

Finally, results from this study indicate that CCI is a feasible and critical approach, given the personalized nature of the crowd data entered. Expert curators were highly involved in merging new milestones with existing ones and in classifying milestones into developmental domains. Figure 12 shows a continuous increase in new information entered by users over time, information that required curation. The experts' work on merging and clarifying milestones directly affected the crowd's adoption of existing milestones and significantly increased the users' chances of viewing statistics for their children's milestones. Our study is consistent with previous research showing that expert facilitators using a dashboard to manage real-time crowd data increased the quality of the information entered [5, 21], and in our case also the information provided.

As with any manual process, there is a cost to the use of human experts in terms of scalability and on-going resources. Enhancing the efficiency of the experts' decision making is needed through the integration of automatic algorithms and assistance using natural language processing (NLP) and machine learning tools, and potentially with a crowd of experts. The manual work done so far by curators provides a natural source of high-quality training and evaluation examples to aid the development of such algorithms. For Baby CROINC's purposes, the curation must be of extremely high quality to maintain user trust. As automatic algorithms are likely to fall short of human quality, the tools will initially assist the curator behind the scenes in providing time-saving top suggestions to curators, which will then be confirmed or overridden. Only when, or if, the tools are assessed as reliable enough to approach perfect performance (as compared to the human gold standard) will we consider replacing some or all of the expert curators' tasks with automatic algorithms. In the meantime, however, as a helpful guide for humans, we expect them to increase the efficiency and scalability of the curators' work.

Limitations

While we believe that the user-driven approach is beneficial and needed in today's online environment, it doesn't allow control of the number and type of milestones entered by parents, thus posing new challenges when seeking to compare different developmental trajectories for children in our system. In addition, showing statistics to parents and particularly the skewness of the entries to early achievement may create a bias in data entry; parents may wish to modify their reported milestone dates to improve their child's statistics and present them as "early achievers," or reluctant to record milestones for which their children are more delayed. Adding information about standardized norms for each milestone can assist parents in interpreting skewed statistics. Characterizing parents who use the system—in terms of their technological proficiency, personal impact, and their user experience—may improve its applicability to diverse socio-economical backgrounds. For research in child development, studies looking at Baby CROINC percentiles' correlation with traditional developmental test scores would be highly informative and would improve our understanding of the value of our system as a pre-screening tool. In absence of a clinical sample of children with known developmental disorders, the validity of Baby CROINC as a pre-screening tool has yet to be demonstrated.

Finally, the amount of expert curation currently grows linearly with system size, making it too costly for massive-scale systems. However, the effectiveness of automatic and curator-assisting algorithms utilizing NLP and machine learning tends to increase as the amount of data grows, and therefore we believe that an algorithm-assisted CCI approach will remain feasible even as the system continues to grow in size.

Potential design improvements

This study raises several ideas for further improvements to Baby CROINC:

1. Integrating NLP and machine learning tools to improve the scalability of the crowd-curating process (either as an aid to the expert and/or to automate aspects of the curation);
2. Prompting parents to add milestones in developmental domains that they have not thought to focus on yet;
3. Presenting age-corrected statistics for children born preterm;
4. Increasing the amount of demographical/background information collected from parents, to improve the personalization of statistics and for child development research;
5. Providing ways to share the developmental conclusions in Baby CROINC with the children's health care providers.

Conclusions

Baby CROINC is a crowd-based platform for child development tracking and screening, which integrates expert-curated crowd data. Its two main design goals are personalization and objectivity, using CCI to resolve the tension between them.

Unlike pre-defined developmental screening tools based on a fixed number and type of milestones, Baby CROINC provides parents with an opportunity to describe development from their perspective, in their own words, and focus on their own interests and concerns. Yet unlike completely free-form narratives, like those on online parenting forums or social media, Baby CROINC provides objective statistical data to educate parents about their child's development.

Our work demonstrates the value of expert involvement in curating crowd-based child development data to provide high-quality, personalized statistics to parents. Further research is important for measuring the impact of using the system on advancing parental knowledge, parenting practices, and for early detection of developmental delays.

Acknowledgements

We thank the team members for their contribution to the Baby CROINC system design and maintenance: Gal Agmon, Tal Bussel, Daniel Moran, Elisheva Rotman Argaman, Ori Sberlo, Yael Schwartz, Gal Shachaf, Danielle Wolinski, Matan Yechieli, Almog Zimelman, and Ohad Zohar. The research of the first author received funding from a Yigal Alon Fellowship; research of the last three co-authors received funding from the European Research Council (under grant agreement 240258) and the Israeli Science Foundation (grant 1501/14).

REFERENCES

1. Paul Beatty. 1995. Understanding the standardized/non-standardized interviewing controversy. *Journal of Official Statistics* 11, 2 (1995), 147.
2. Ayelet Ben-Sasson and Alice S Carter. 2012. The application of the first year inventory for ASD screening in Israel. *Journal of Autism and Developmental Disorders* 42, 9 (2012), 1906–1916.
3. Ayelet Ben-Sasson and Elad Yom-Tov. 2016. Online Concerns of Parents Suspecting Autism Spectrum Disorder in Their Child: Content Analysis of Signs and Automated Prediction of Risk. *J Med Internet Res* 18, 11 (22 Nov 2016), e300. DOI: http://dx.doi.org/10.2196/jmir.5439
4. Jay M Bernhardt and Elizabeth M Felter. 2004. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. *Journal of Medical Internet Research* 6, 1 (2004), e7.
5. Joel Chan, Steven Dang, and Steven P. Dow. 2016. Improving Crowd Innovation with Expert Facilitation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1223–1235. DOI: http://dx.doi.org/10.1145/2818048.2820023
6. Geraldine Dawson, Sharon B Ashman, and Leslie J Carver. 2000. The role of early experience in shaping behavioral and brain development and its implications for social policy. *Development and Psychopathology* 12, 04 (2000), 695–712.
7. Deborah Dobrez, Anthony Lo Sasso, Jane Holl, Madeleine Shalowitz, Scott Leon, and Peter Budetti. 2001. Estimating the cost of developmental and behavioral screening of preschool children in general pediatric practice. *Pediatrics* 108, 4 (2001), 913–922.
8. Mia Wechsler Doron, Emma Trenti-Paroli, and Dana Wechsler Linden. 2013. Supporting parents in the NICU: A new app from the US, 'MyPreemie': A tool to provide parents of premature babies with support, empowerment, education and participation in their infant's care. *Journal of Neonatal Nursing* 19, 6 (2013), 303–307.
9. Kate Ellis-Davies, Elena Sakkalou, Nia C Fowler, Elma E Hilbrink, and Merideth Gattis. 2012. CUE: The continuous unified electronic diary method. *Behavior Research Methods* 44, 4 (2012), 1063–1078.
10. C. Ferre, W. Callaghan, C. Olson, A. Sharma, and W. Barfield. 2016. Effects of Maternal Age and Age-Specific Preterm Birth Rates on Overall Preterm Birth Rates - United States, 2007 and 2014. *MMWR Morb. Mortal. Wkly. Rep.* 65, 43 (Nov 2016), 1181–1184.
11. Jeana Frost and Michael Massagli. 2008. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one anotherâĂŹs data. *Journal of Medical Internet Research* 10, 3 (2008), e15.
12. Joseph Hagan. 2008. *Bright futures : guidelines for health supervision of infants, children, and adolescents*. American Academy of Pediatrics, Elk Grove Village, IL.
13. Mary L Hediger, Mary D Overpeck, W Ruan, and James F Troendle. 2002. Birthweight and gestational age effects on motor and social development. *Paediatric and Perinatal Epidemiology* 16, 1 (2002), 33–46.
14. Voxiva Inc. 2017. Text4Baby™. (2017). https://www.text4baby.org/
15. Juyoung Jang, Jodi Dworkin, and Jessie Connell. 2012. Babycenter. com: New parent behavior in an online community. (2012).
16. Thomas Kellaghan. 1982. *The Effects of Standardized Testing*. Springer Netherlands, Dordrecht.
17. Kaylyn Khoo, Penny Bolt, Franz E Babl, Susan Jury, and Ran D Goldman. 2008. Health information seeking by parents in the Internet age. *Journal of Paediatrics and Child Health* 44, 7-8 (2008), 419–423.
18. Julie A. Kientz, Rosa I. Arriaga, and Gregory D. Abowd. 2009. Baby Steps: Evaluation of a System to Support Record-keeping for Parents of Young Children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1713–1722. DOI: http://dx.doi.org/10.1145/1518701.1518965
19. Ana M Leon, Shannon Holliker, and Julie Pepe. 2015. The importance of the first 5 years: Pediatrician identification of developmental delays and other related concerns. *Journal of Social Service Research* 41, 4 (2015), 425–444.
20. Ben MacNeill. 2017. Trixie Tracker™. (2017). https://www.trixietracker.com/
21. David Merritt, Jasmine Jones, Mark S. Ackerman, and Walter S. Lasecki. 2017. Kurator: Using The Crowd to Help Families With Personal Curation Tasks. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1835–1849. DOI: http://dx.doi.org/10.1145/2998181.2998358
22. Janet M Morahan-Martin. 2004. How internet users find, evaluate, and use online health information: A cross-cultural review. *CyberPsychology & Behavior* 7, 5 (2004), 497–510.
23. Committee on Children with Disabilities and others. 2001. Developmental surveillance and screening of infants and young children. *Pediatrics* 108, 1 (2001), 192–195.
24. Sally Ozonoff, Gregory S Young, Mary Beth Steinfeld, Monique M Hill, Ian Cook, Ted Hutman, Suzanne Macari, Sally J Rogers, and Marian Sigman. 2009. How early do parent concerns predict later autism diagnosis? *Journal of Developmental and Behavioral Pediatrics: JDBP* 30, 5 (2009), 367.
25. Lars Plantin and Kristian Daneback. 2009. Parenthood information and support on the internet: A literature review of research on parents and professionals online. *BMC Family Practice* 10, 34 (2009), 1–12.
26. Peter Sacks. 1999. *Standardized minds : the high price of America's testing culture and what we can do to change it*. Perseus Books, Cambridge, Mass.
27. Steven Shelov. 2009. *Caring for your baby and young child : birth to age 5*. Bantam, New York.
28. Jane Squires. 2009. *Ages & stages questionnaires : a parent-completed child monitoring system*. Paul H. Brookes Publishing Company, Baltimore, Maryland.
29. Hyewon Suh, John R. Porter, Alexis Hiniker, and Julie A. Kientz. 2014. @BabySteps: Design and Evaluation of a System for Using Twitter for Tracking Children's Developmental Milestones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2279–2288. DOI: http://dx.doi.org/10.1145/2556288.2557386
30. Brynn K Wainstein, Katy Sterling-Levis, Sally A Baker, Jonathan Taitz, and Michael Brydon. 2006. Use of the Internet by parents of paediatric patients. *Journal of Paediatrics and Child Health* 42 (2006), 528–532.
31. Paul Wicks, Michael Massagli, Jeana Frost, Catherine Brownstein, Sally Okun, Timothy Vaughan, Richard Bradley, and James Heywood. 2010. Sharing health data for better outcomes on PatientsLikeMe. *Journal of Medical Internet Research* 12, 2 (2010), e19.