JAMA Pediatrics | Original Investigation | IMPACT OF POLICY ON CHILDREN

Data Collection Practices of Mobile Applications Played by Preschool-Aged Children

Fangwei Zhao, BA; Serge Egelman, PhD; Heidi M. Weeks, PhD; Niko Kaciroti, PhD; Alison L. Miller, PhD; Jenny S. Radesky, MD

IMPORTANCE Child-directed mobile applications (apps) have been found to collect digital identifiers and transmit them to third-party companies, a potential violation of federal privacy rules. This study seeks to examine the differences in app data collection and sharing practices by evaluating the sociodemographic characteristics of the children who play them.

OBJECTIVE To examine data collection and sharing practices of 451 apps played by young children and to test associations with child sociodemographic characteristics.

DESIGN, SETTING, AND PARTICIPANTS This study used data from the baseline phase of the Preschooler Tablet Study, a prospective cohort study conducted from August 2018 to January 2020. This study used a population-based sample. A convenience sample of the parents of preschool-aged children was recruited from pediatric offices, childcare centers, social media posts, and an online participant registry. Eligibility criteria included (1) parent or guardian of a child aged 3 to 5 years, (2) parent or guardian who lived with the child at least 5 days per week, (3) participants who spoke English, and (4) a child who used an Android (Google LLC) device. All interactions with participants were through email, online surveys, and mobile device sampling.

EXPOSURES Sociodemographic characteristics were assessed by parental report.

MAIN OUTCOMES AND MEASURES This study tested the hypothesis that data transmissions to third-party domains are more common in apps played by children from low-socioeconomic-status homes. Child app usage was assessed via a mobile sampling app for an average of 9 days. Persistent identifier data transmissions to third-party domains were quantified for each app using an instrumented Android environment with monitoring of network traffic; for each child, the counts of total data transmissions were calculated, and the total third-party domains were detected for the apps they played.

RESULTS Our sample comprised 124 children who used Android devices (35 tablets, 89 smartphones; 65 girls [52%]; mean [SD] age, 3.85 [0.57] years; 87 non-Hispanic White [71%]). One hundred twenty of participating parents (97%) were women. Of 451 apps tested, 303 (67%) transmitted persistent identifiers to 1 to 33 third-party domains. Child data transmission counts ranged from 0 to 614 (median [interquartile range], 5.0 [1-17.5]) and third-party domain counts from 0 to 399 (4.0 [1-12.5]). In multivariable negative binomial regression models, higher transmission and third-party domain rates per app were positively associated with older age (rate ratio, 1.67 [95% CI, 1.20-2.33]; $P = .002$ and 1.69 [95% CI, 1.26-2.27]; $P < .001$, respectively) and lower parent educational attainment (eg, high school or General Educational Development or less rate ratio, 2.29 [95% CI, 1.20-4.39]; $P = .003$ and 2.05 [95% CI, 1.13-3.70]; $P < .02$, respectively), but not with household income.

CONCLUSIONS AND RELEVANCE This study found that apps used by young children had a high frequency of persistent identifier transmissions to third-party companies, suggesting that federal privacy rules are not being enforced. Older children, those with their own devices, or those from lower-education households may be at higher risk of potential privacy violations.

JAMA Pediatr. 2020;174(12):e203345. doi:10.1001/jamapediatrics.2020.3345
Published online September 8, 2020.

+ Supplemental content

+ Editorial

Author Affiliations: Department of Pediatrics, University of Michigan Medical School, Ann Arbor (Zhao, Radesky); Usable Security and Privacy Group, International Computer Science Institute, Berkeley, California (Egelman); Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor (Weeks); Center for Human Growth and Development, University of Michigan, Ann Arbor (Kaciroti, Radesky); Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor (Kaciroti, Miller); Department of Health Behavior and Health Education, University of Michigan School of Public Health, Ann Arbor (Miller).

Corresponding Author: Jenny S. Radesky, MD, Developmental Behavioral Pediatrics, University of Michigan Medical School, 300 N Ingalls St, #1107, Ann Arbor, MI 48109 (jradesky@med.umich.edu).

Over the past decade, children's use of mobile devices such as smartphones and tablets has become almost universal,[1] and many children have their own tablet.[2] Of the 2.2 million applications (apps) on the Apple App Store and 2.7 million apps on the Google Play store, tens of thousands are games or educational apps marketed to children.[3] Users generate digital footprints while using mobile apps and internet-connected devices—comprising data such as websites visited, gameplay behavior, purchases made, or other identifiers hardwired into mobile devices (eg, device serial number). These data are often shared with third-party companies for the purposes of usage analytics and creation of digital profiles that track user behavior across digital services for marketing and other purposes.

Under the Children's Online Privacy Protection Act (COPPA),[4] platforms and creators of digital products must obtain verifiable parental consent before collecting and sharing personally identifiable data (ie, anything traceable back to the child, such as location, email, or device identifier) from children younger than 13 years for behavioral advertising purposes. However, enforcement of COPPA has been mostly limited to actions filed against large platforms such as TikTok[5] and YouTube.[6] Because it is not apparent to users when digital data are collected or transmitted to third-party companies, it is difficult for consumers to make informed choices about which apps to install for their children. In prior research, parents and children reported not understanding digital privacy concepts, including apps' data collection practices,[7,8] targeted advertising, or the storage of their private information.[9] Given poorly enforced child privacy regulations and opaque data collection practices not understood by consumers, it is important to examine potential privacy violations in children's apps. Young children (5 years and younger) deserve particular focus because of the large quantity of child-directed apps, many of which are monetized through in-app purchases and advertisements.[10]

Two prior studies have examined data collection and sharing from apps marketed to children on the Google Play app store. Binns and colleagues[11] used static app analysis (ie, analyzing app source code to find code that directs data collection to third parties) on 959 000 apps from the US and UK Google Play stores. They found that apps targeting children had among the highest number of third-party trackers. Reyes et al[12] used dynamic analysis to track the data transmissions from 5855 of the most popular free Android children's apps and showed that the majority had potential COPPA violations. We aimed to build on these studies by examining the data collection practices of apps played by children in a cohort study and quantify whether such practices are associated with user sociodemographic characteristics. Prior research suggests that parents with higher income are more likely to monitor the apps their children download[1] and have more digital privacy knowledge[13] and concerns.[14] Therefore, it is possible that children from low socioeconomic strata may have higher rates of digital privacy violations. We hypothesized that data transmissions to third-party domains would be more common in apps played by children from low socioeconomic status (ie, lower household income, lower parent educational attainment).

Key Points

Question What are the data collection and sharing practices of mobile applications (apps) played by young children, and do potential privacy violations differ by child characteristics?

Findings Two-thirds of apps played by 124 preschool-aged children in this cohort study showed collection and sharing of persistent digital identifiers. Children who were older, had their own mobile devices, played a higher number of apps, or were from lower-education households had higher counts of data transmissions to a higher number of third-party domains, whereas only 8% of children played apps that showed zero identifier transmissions.

Meaning Digital identifiers collected from children's mobile devices can be used for profiling and marketing purposes; this study's results suggest that potential violations of child digital privacy laws are common, and enforcement of the Children's Online Privacy Protection Act is needed.

Methods

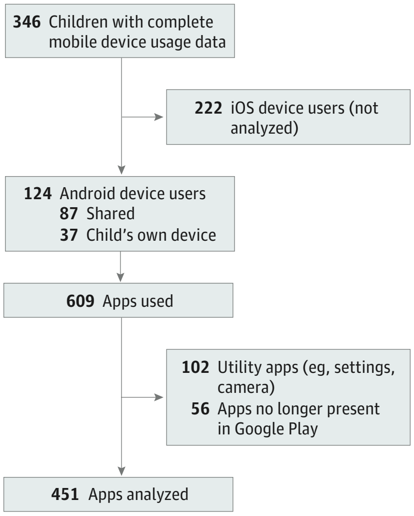
Study Design

In this cohort study, we analyzed data from the first wave of the Preschooler Tablet Study, a longitudinal study of child mobile media use, development, and behavior (National Institute of Child Health and Human Development No. R21HD094051). In this study, caregivers of preschool-aged children completed online questionnaires and were asked to provide a 9-day sample of their child's mobile device usage data at baseline, 3 months, and 6 months. Data from the baseline data collection wave (August 2018 to May 2019) are included in the present manuscript. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline, and it was approved by the institutional review board of the University of Michigan. Parents completed written online informed consent for themselves and their child; a waiver of documentation was allowed because of the minimal risk of the study.

Participants

Parents and children were recruited through flyers in pediatric offices and the community, as well as through online participant registry- and social media–based recruitment. Eligibility criteria included (1) parents who were legal guardians of a child aged 3 to 5 years, (2) parents who lived with the child at least 5 days per week, (3) parents who understood English sufficiently to complete questionnaires and provide consent, and (4) families that owned at least 1 Android or iOS (Apple Inc) tablet or smartphone. Exclusion criteria included the presence of child developmental delays or use of psychotropic medication. Of the 346 children with mobile device data, we only included children who used Android devices (n = 124, see **Figure 1**) because our transmission analysis utilizes an Android framework. Android users were more likely to be from lower-income households (mean [SD] income-to-needs ratio, 2.17 [1.60] vs 3.48 [1.61]; Wilcoxon 2-sample test, $P < .001$) and have parents without a 4-year college degree (65 of 124

Figure 1. CONSORT Diagram: Participants Included and Apps Analyzed in the Present Study


346 Children with complete mobile device usage data
222 iOS device users (not analyzed)
124 Android device users
87 Shared
37 Child's own device
609 Apps used
102 Utility apps (eg, settings, camera)
56 Apps no longer present in Google Play
451 Apps analyzed

[52.4%] vs 64 of 220 [29.1%]; χ^2 test, $P < .001$), compared with iOS users.

Survey Measures

After consenting to participate in the study, parents were sent an email message with a link to a Research Electronic Data Capture, or REDCap[15] survey (Vanderbilt University), in which they reported their child's age, sex, preschool or childcare enrollment, and race/ethnicity (options provided by the study team based on prior evidence that non-Hispanic White race/ethnicity is associated with less media use[1]); parents' age, sex, educational attainment, marital status, and employment status; and household income and size (from which we calculated income-to-needs ratio).

Child Mobile App Use

Parents were also sent instructions via email for mobile device sampling, a novel objective assessment method described in detail elsewhere.[16] Participants with smartphones running Android operating systems (eg, Samsung, Motorola) were instructed on how to install a passive sensing app, Chronicle (OpenLattice), which queries the Google app usage statistics manager to provide a continuous log of the start and end times of every app used. The log is automatically uploaded to secure cloud-based storage, from which we downloaded a data file after approximately 9 days (mean [range], 8.96 [4-12] days) and instructed the parent to uninstall the app. For children with their own mobile device (n = 37), a list of all apps accessed during the sampling week was generated from the data file. For children who shared a mobile device (ie, used their parent's phone or shared a tablet with siblings, n = 87), we asked parents to indicate which apps their child used during the sampling week so that we could omit apps used by parents or siblings from the data file.

From these data files, we generated a list of all Android apps used by children, comprising 609 unique app package names. Of these, 102 were utility apps and 56 were no longer present in the Google Play app store, so 451 apps were analyzed for data transmission practices (Figure 1).

Privacy Behavior Analysis

We analyzed transmissions from apps based on the method from Reyes et al.[12] Apps were downloaded from the Google Play store to an instrumented network of Android devices. This means that the devices were running a modified version of the Android operating system that inspects all network traffic (ie, flows of data to and from the device) generated by the app being tested. Apps were played using Android's Application Exerciser Monkey, which automates the execution of apps by simulating user inputs in a pseudorandom manner. Each app was scheduled to run for 10 minutes, during which time logs were generated about data transmissions. To our knowledge, this method is the only existing approach for examining data transmissions from large volumes of apps, and it has been validated against actual user interaction with apps.[12] Because the exerciser monkey cannot follow visual cues, it cannot know when it needs to click a button—eg, to dismiss a dialog box. This might result in not all possible paths being explored by the monkey; therefore, the results represent a lower bound of what an app can do while interacting with a human user. In other words, the exerciser monkey may incur false negatives but does not generate false positives. Therefore, we performed multiple bouts of data collection between August 1, 2019, and November 1, 2019, to reduce the number of false-negative results.

Postprocessing of network flows involves searching for identifiers as string values associated with the particular testing device, such as the phone number, location coordinates, Android ID, and advertising ID (see **Table 1** for list and definitions), either verbatim or after the app had applied various obfuscations (eg, character encodings, cryptographic hashing). If identifiers were found in an outgoing flow from the tested app, the internet protocol address and domain name of the destination were recorded.

For each app, we calculated 2 variables: the number of unique data transmissions detected and the number of unique third-party domains to which transmissions occurred. We excluded transmissions with data types that are not obvious identifiers by themselves. If data transmissions were detected for an app at 1 time point but not any others, we assigned that app the data transmission count from the time point at which transmissions were detected. Apps with transmissions detected at more than 1 time point were assigned the higher number of transmissions. Apps for which no data transmissions were ever detected were assigned a count of 0.

We then merged app transmission information with child-level data regarding which apps each child played during the sampling week. For each child, based on their app list, we calculated (1) the total data transmission count (sum of transmissions for apps played by child) and (2) the third-party domain count (sum of the number of unique domains).

Statistical Analysis

We calculated descriptive statistics on data transmission and domain counts for each app and compared whether apps with

Table 1. Data Types Transmitted From 451 Android Apps Played by 124 Preschool-Aged Children

Data type	Definition	Transmission frequency, No.
Advertising ID	A unique, user-resettable ID for advertising, provided by Google Play services	2283
Android ID	A unique ID to each device. It is used to identify each device for market downloads and specific gaming applications	824
Hardware ID (device serial No.)	A string that uniquely and consistently identifies a given specific machine or device	37
Geolocation (latitude and longitude)	Allows identification or estimation of the real-world geographic location of an object	30
IMEI	A unique number to identify specific models of mobile phones (eg, satellite phones)	19
Router SSID	The name of the wireless network to which the device is connected	19
Router MAC address	A unique 12-character hexadecimal number (for example, 00:00:00:00:00:0X) that identifies the router to which the device is connected. Since routers remain in fixed places, the MAC address is often used as a proxy for location data	12
Wi-Fi MAC address	A unique 12-character hexadecimal number (for example, 00:00:00:00:00:0X) that a device uses to identify itself on a network	11
Phone	User's phone number	10
Email	User's email address	5
Real name	User's real name	5
GSF ID	A permanent, unique 16-character hexadecimal number requested by a device the first time a Google account is logged in	4
SIM ID	Users' SIM card's 19-digit identification number	1

Abbreviations: GSF indicates Google Services Framework; ID, identifier; IMEI, International Mobile Equipment Identity; MAC, media access control; SIM, subscriber identification module; SSID, network name.

the Designed for Families (DFF) designation (ie, categorized by Google as being specifically directed to children) or those with highest popularity in our study population (top decile) showed different digital privacy practices. We then examined bivariate associations between child data transmission count, third-party domain count, and sociodemographic characteristics. Bivariate analyses used Spearman correlations, Wilcoxon 2-sample test, Kruskal-Wallis test, or Fisher exact test as appropriate. Because transmission and domain counts were overdispersed, we conducted a negative binomial regression with a log link function, including an offset variable for the number of apps the child played that underwent privacy behavior analysis. We built multivariable models by including all independent variables showing marginal (*P* < .20) bivariate associations with data transmission variables and then reducing the models manually to only include independent variables with 2-sided *P* values <.05. We reported transmission and domain rate ratios per app for each independent variable from the multivariable model. All data analyses were conducted from January 10, 2020, to April 15, 2020, using R version 3.5.2 (R Foundation) or SAS version 9.4 (SAS Institute).

Results

App-Level Analysis

Of the 451 apps examined, 303 (67%) showed transmission of identifiers to third-party domains. Advertising IDs (2283 transmissions), Android IDs (824 transmissions), hardware IDs (ie, device serial number; 37 transmissions), and geolocation (ie, Global Positioning System coordinates; 30 transmissions) were the most commonly transmitted data types (Table 1). The median number of transmissions was 3.0 (interquartile range [IQR], 0-10), with a maximum of 57 (*Happy Glass,* played by 4 children, transmitted advertising IDs 50 times and Android IDs 7 times, to 33 different domains). The median number of data types per app was 2.0 (IQR, 0-2; maximum 4) and median domains per app was 2.0 (IQR, 0-6; maximum 33). Transmission and domain counts for all apps tested are shown in eTable 1 in the Supplement.

Apps in the DFF program (n = 153) had fewer data transmissions to fewer domains compared with apps not in DFF (eTable 2 in the Supplement) and were less likely to transmit advertising IDs (46% vs 69%; *P* < .001), Android IDs (44% vs 56%; *P* = .02), and geolocation (0% vs 4%; *P* = .02). Of DFF apps, 68 (44%) had zero transmissions, compared with 80 (27%) non-DFF apps (*P* < .001). The most popular apps in the study population (top 10%; played by ≥3 children) showed fewer data transmissions, domain counts, and unique data types than the lower 90% (eTable 2 in the Supplement).

Child-Level Analysis

Table 2 shows the characteristics of the 124 children (mean [SD] age, 3.85 [0.57] years; 65 girls [52%]) included in this analysis. Children were primarily non-Hispanic White (87 [71%]), showed a wide distribution of parent education (high school, General Education Development, or less, 15 [12%]; some college, 50 [40%]; 4-year college degree, 31 [25%]; more than 4-year college degree, 28 [23%]) and income-to-needs ratio (mean [SD] range, 2.17 [1.60] 0.14-6.02), and most were enrolled in center-based childcare or preschool. Almost one-third of children had their own mobile device (37 [30%]); the remainder shared with a parent or sibling (87 [70%]). Children's data transmission and third-party domain counts shared a similar distribution (**Figure 2**).

In bivariate analyses, most sociodemographic characteristics were not associated with data transmission counts or third-party domain counts (child sex, parent age, parent marital status, and income-to-needs ratio) (Table 2). Children with their own mobile device showed higher transmission counts (median [IQR], 18.0 [5.0-142.0]) and domain counts (median [IQR], 11.0 [4.0-99.0]; *P* < .001) compared with children who shared devices with other family members (median [IQR], 3.0 [1.0-9.0] transmission counts and 3.0 [1.0-7.0] domain counts; *P* < .001). However, this was due to children with their own devices playing a higher number of apps and was no longer significant in regression models with offset for number of apps analyzed. In multivariable regression models (**Table 3**), higher transmission and domain rate ratios per app were associated

Table 2. Sociodemographic Characteristics of 124 Preschool-Aged Children Participating in the Study Population and Bivariate Associations With Transmission and Domain Counts

		Median (IQR) or Spearman ρ	
Characteristic	No. (%)[a]	Transmission counts	Domain counts
Child sex			
Female	65 (52)	5.0 (1.0-14.0)	4.0 (1.0-11.0)
Male	59 (48)	6.0 (1.0-21.0)	4.0 (1.0-14.0)
Child age, mean (SD) [range], y	3.85 (0.57) [3.01-4.99]	0.26[b]	0.27[b]
Child race/ethnicity			
Non-Hispanic White	87 (71)	5.0 (1.0-17.0)	4.0 (1.0-12.0)
Other[c]	35 (29)	6.0 (1.0-31.0)	4.0 (1.0-20.0)
Parent sex			
Female	120 (97)	5.5 (1.0-17.5)	4.0 (1.0-12.5)
Male	4 (3)	3.5 (1.0-249.5)	3.0 (1.0-170.5)
Parent age, mean (SD) [range], y	33.74 (4.8) [24.0-47.1]	−0.08	−0.08
Parent marital status			
Married or partnered	106 (86)	5.0 (1.0-18.0)	4.0 (1.0-13.0)
Single, divorced, or separated	18 (14)	11.0 (1.0-17.0)	7.0 (1.0-10.0)
Parent educational attainment			
High school/GED or less	15 (12)	4.0 (1.0-61.0)[b]	3.0 (1.0-40.0)[d]
Some college or 2-y degree	50 (40)	10.0 (2.0-27.0)[b]	6.0 (1.0-19.0)[d]
4-y College degree	31 (25)	4.0 (1.0-28.0)[b]	3.0 (1.0-13.0)[d]
More than 4-y college degree	28 (23)	2.0 (1.0-6.0)[b]	2.0 (1.0-4.5)[d]
Parent employment status			
None	43 (35)	6.0 (1.0-21.0)	5.0 (1.0-14.0)
1 Full-time job	51 (41)	6.0 (1.0-18.0)	4.0 (1.0-13.0)
1 Part-time job	22 (18)	2.0 (1.0-28.0)	2.0 (1.0-20.0)
Multiple jobs	8 (7)	1.0 (1.0-7.0)	1.0 (1.0-5.5)
Child school or childcare attendance			
Center-based program	76 (62)	5.0 (1.0-20.0)	4.0 (1.0-13.5)
Home-based program	9 (7)	6.0 (1.0-15.0)	5.0 (1.0-12.0)
Stays home with parent or caregiver	38 (31)	5.0 (1.0-14.0)	3.5 (1.0-11.0)
Only child			
Yes	17 (14)	5.0 (1.0-17.0)	4.0 (1.0-11.0)
No	107 (86)	6.0 (3.0-31.0)	4.0 (2.0-19.0)
Income-to-needs ratio, mean (SD) [range]	2.17 (1.60) [0.14-6.02]	−0.12[e]	−0.11
Child app variables			
Shared mobile device			
No	37 (30)	18.0 (5.0-142.0)[f]	11.0 (4.0-99.0)[f]
Yes	87 (70)	3.0 (1.0-9.0)[f]	3.0 (1.0-7.0)[f]
No. of apps, median (IQR) [range]	3.0 (2-12)	0.70[e]	0.69[e]
Played	3.0 (2-12) [1-94]	0.70[f]	0.69[f]
Analyzed	3.0 (2-6) [1-77]	0.77[f]	0.76[f]
Total data transmission count, median (IQR) [range]	5.0 (1-17.5) [0-614]	NA	NA
Third-party domain count, median (IQR) [range]	4.0 (1-12.5) [0-399]	NA	NA

Abbreviations: IQR indicates interquartile range; GED, General Educational Development; NA, not applicable.

[a] Values are expressed as No. (%) unless otherwise specified.

[b] Bivariate analysis *P* value <.01.

[c] Other constitutes Black, Hispanic, Asian, and multiple races.

[d] Bivariate analysis *P* < .05.

[e] Bivariate analysis *P* < .20.

[f] Bivariate analysis *P* < .001.

with older child age (rate ratio, 1.67 [95% CI, 1.20-2.33]; *P* = .002 and 1.69 [95% CI, 1.26-2.27]; *P* < .001) and lower parent education (high school or General Educational Development, or less, 2.29 [95% CI, 1.20-4.39]; *P* = .003 and 2.05 [95% CI, 1.13-3.70]; *P* < .02).

Discussion

By combining dynamic app transmission analysis with child-level data from a population-based cohort, this study was able

Figure 2. Percentage of 124 Preschool-Aged Children With Designated Data Transmission Counts and Third-Party Domain Counts

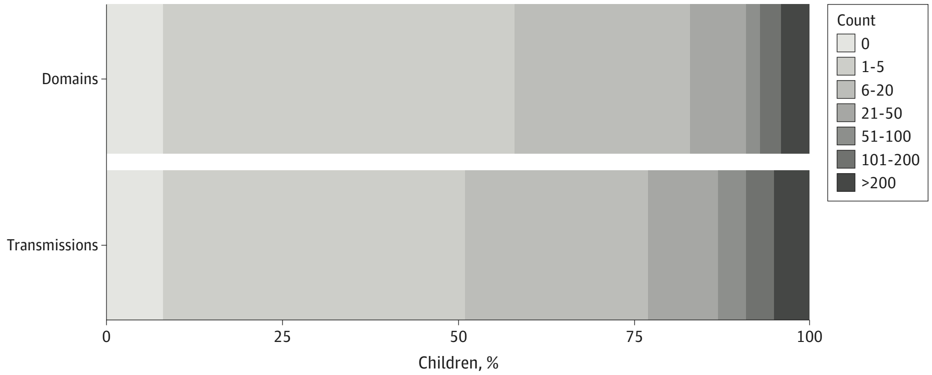


Table 3. Associations Between Child Sociodemographic and Media Use Characteristics With Data Transmission Rates and Domain Rates per App

	Rate ratio (95% CI)[a]	
Characteristic	Transmission	Domain
Child age, y	1.67 (1.20-2.33)	1.69 (1.26-2.27)
Shared mobile device		
No	NA	NA
Yes	NA	NA
Parent education		
High school, GED, or less	2.29 (1.20-4.39)	2.05 (1.13-3.70)
Some college or 2-y degree	2.21 (1.34-3.65)	1.95 (1.23-3.10)
4-y College degree	2.95 (1.69-5.12)	2.10 (1.26-3.50)
More than 4-y degree	1 [Reference]	1 [Reference]
Household income-to-needs ratio	NA	NA

Abbreviations: GED, General Educational Development; NA, not applicable.

[a] Rate ratios are exponentiated coefficients that represent the transmission rate ratio between groups whose covariate values differ by 1 unit.

to examine sociodemographic associations of potential digital privacy violations. Our results suggest a high prevalence of digital identifier transmission from apps played by preschool-aged children. Children raised by parents without advanced degrees showed 2 to 3 times higher rates of identifier transmissions to third-party domains. Older children also showed a significantly higher number of data transmissions to more third-party domains. These associations were partially explained by the fact that these children played a higher number of apps, but they may also be playing apps with less age-appropriate design or more data trackers.

Two-thirds of the apps we tested showed transmission of identifiers to third-party domains. The most commonly transmitted identifier was the advertising ID, which is used to create advertising behavioral profiles of users. Although the advertising ID is technically resettable, this requires technical knowledge that children (and their parents) likely do not have. Other persistent identifiers traceable back to the user include the Android ID, email address, Wi-Fi or router information (which can be used to identify location), and geolocation data. These types of data transmissions were more common in apps used by children with their own mobile devices due to the higher number of apps installed. In this cohort, children with their own mobile device also played a higher number of general audience apps (eg, *Subway Surfers, Color Road*), which may not comply with COPPA. However, it is also possible that parents who allow their child to use their mobile device are not aware of when their child uses general audience apps.

Our findings build upon prior large-scale static[11] and dynamic[12] app analyses showing widespread collection and sharing of device-based identifiers by apps marketed to children. In addition, our results suggest that general audience apps need to consider their child audiences. For example, some children in our study used apps that transmit geolocation data, such as the *McDonald's* app, and games such as *hole.io* and *SpeedBall*. Children may easily download general audience apps from Google Play when parental controls are not enabled. It is also possible that children install adult-directed apps through advertisements that appear in children's apps,[10] where they can easily be clicked and installed.

Understanding the extent of digital data collection from children is important for several reasons. Children's privacy is listed under the United Nations Convention on the Rights of the Child and is recognized within digital contexts by policy experts[17] and the American Academy of Pediatrics.[18] Although it may be more intuitive to parents to protect their child's location and contact information data, there are other aspects of digital privacy. Digital profiles constructed through patterns of digital behaviors linked by device identifiers may identify child vulnerabilities that could be exploited[8]—such as impulsive gameplay behavior or demographic inferences—that could be used to advertise unhealthy products or encourage more purchases. These vulnerabilities may be more difficult to protect when parents have lower educational attainment and digital literacy.[13,14]

App developers and platform designers play an important role in reducing or eliminating digital identifier collection (ie, data minimization) because the types of persistent identifiers that track users across apps are not needed for the types of analytics that help apps function better. For example, PBS KIDS apps create a novel hashed identifier for each user that cannot be traced back to their identity or across other apps but allows tracking of app functioning and use (interview with PBS KIDS staff, September 2019). The Google Play store has recently strengthened privacy guidelines for apps in the DFF section (eg, not collecting location data), but data minimization standards have not been implemented. More than half of the DFF apps in our analysis collected and shared device identifiers.

Without design and regulatory frameworks that make data minimization and digital privacy the default, parents must continue to act as gatekeepers of their children's online privacy.[7] However, adults often have inaccurate understandings of the internet[19] and online data flows.[9] Although many parents express concern about their children's digital privacy, most have limited knowledge of privacy practices, such as those described by privacy policies,[7] which can be difficult to decipher.[20] Until data collection and sharing practices are transparent, parents may wish to install apps from trusted developers (eg, PBS KIDS), learn more about digital privacy from resources such as Common Sense Media, or look up the data collection practices of apps on databases such as http://search.appcensus.io.[12]

Limitations

This study has several limitations. For children with shared devices, we relied on parents reporting which apps their child typically uses, which may have been inaccurate. In addition, we only examined data collection practices of apps played on Android devices. Our results cannot comment on the privacy behaviors of apps on the Apple App Store, which has recently strengthened its requirements regarding privacy in children's apps[21] but only make up about 20% of the mobile platform market share. Although we conducted multiple runs to detect data transmissions, we may have underestimated the true extent of identifier transmission because our method might not imitate all user interactions. In addition, some transmissions were to analytic companies, such as Crashlytics, and therefore may not be used for behavioral advertising purposes; however, the most common third-party domains in this study provided advertising-related services (eTable 3 in the Supplement). Because of frequent changes to app code, developers, and privacy policies, app analyses are a moving target, which is a general limitation of app analysis research.

Conclusions

This study examined data collection and sharing practices of apps played by preschool-aged children and tested differences by child characteristics. Our findings suggest that the collection and sharing of children's data are highly prevalent, and disparities exist by parent education. These results highlight the need for comprehensive testing of app and platform data collection practices by regulatory bodies so that updated privacy legislation can be crafted that adequately protects children's rights in the modern digital environment.

ARTICLE INFORMATION

Accepted for Publication: May 21, 2020.

Published Online: September 8, 2020. doi:10.1001/jamapediatrics.2020.3345

Author Contributions: Dr Radesky had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Zhao, Radesky.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Zhao, Radesky.
Critical revision of the manuscript for important intellectual content: Egelman, Weeks, Kaciroti, Miller, Radesky.
Statistical analysis: Zhao, Weeks, Kaciroti.
Obtained funding: Radesky.
Administrative, technical, or material support: Egelman.
Supervision: Miller, Radesky.

Conflict of Interest Disclosures: Dr Egelman reported that data regarding mobile app behaviors was provided by AppCensus Inc, which is a startup from prior research. Dr Egelman reported serving as Chief Technology Officer for AppCensus Inc and is currently a shareholder. Dr Weeks reported receiving grants from the National Institute of Child Health and Human Development outside the submitted work. Dr. Kaciroti reported receiving grants from the National Institutes of Health outside the submitted work. Dr Radesky reported receiving grants from the National Institute of Child Health and Human Development, personal fees from Melissa & Doug LLC, personal fees from Public Broadcasting Service, and research funding from Common Sense Media, which is recommended as a parent resource in this manuscript, all outside the submitted work. No other disclosures were reported.

Funding/Support: This study was supported by grant 1R21HD094051 from the Eunice Kennedy Shriver National Institute of Child Health and Development (Dr Radesky) and grant CNS-1817248 from the National Science Foundation (Dr Egelman). Research Electronic Data Capture (REDCap), and recruitment support was provided by grant UL1TR002240 from the Michigan Institute for Clinical & Health Research (University of Michigan).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Rideout V. *The Common Sense Census: Media Use by Kids Age Zero to Eight.* Common Sense Media Inc; 2017.

2. Kabali HK, Irigoyen MM, Nunez-Davis R, et al. Exposure and use of mobile media devices by young children. *Pediatrics*. 2015;136(6):1044-1050. doi:10.1542/peds.2015-2151

3. Vaala S, Ly A, Levine MH. *Getting a Read on the App Stores: A Market Scan and Analysis of Children's Literacy Apps.* Joan Ganz Cooney Center at Sesame Workshop; 2015.

4. United States Federal Trade Commission. Complying with COPPA: frequently asked questions. Accessed October 1, 2019. https://www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequently-asked-questions

5. United States Federal Trade Commission. Video social networking app Musical.ly aggress to settle FTC allegations that it violated children's privacy law. Published February 27, 2019. Accessed October 1, 2019. https://www.ftc.gov/news-events/press-releases/2019/02/video-social-networking-app-musically-agrees-settle-ftc

6. United States Federal Trade Commission. Google and YouTube will pay record $170 million for alleged violations of children's privacy law. Published September 4, 2019. Accessed October 1, 2019. https://www.ftc.gov/news-events/press-releases/2019/09/google-youtube-will-pay-record-170-million-alleged-violations

7. Zhao J. Are children well-supported by their parents concerning online privacy risks, and who supports the parents? Published April 28, 2018. Accessed September 1, 2019. https://arxiv.org/pdf/1809.10944.pdf

8. Livingstone S, Stoilova M, Nandagiri R. *Children's Data and Privacy Online: Growing Up in a Digital Age*. London School of Economics and Political Science; 2019.

9. Yao Y, Lo Re D, Wang Y. Folk models of online behavioral advertising. Paper presented at: 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing; February 1-March 25, 2017; Portland, Oregon. Accessed September 1, 2019. https://www.researchgate.net/publication/313738097_Folk_Models_of_Online_Behavioral_Advertising

10. Meyer M, Adkins V, Yuan N, Weeks HM, Chang Y-J, Radesky J. Advertising in young children's apps:

a content analysis. *J Dev Behav Pediatr*. 2019;40(1):32-39. doi:10.1097/DBP.0000000000000622

11. Binns R, Lyngs U, Van Kleek M, Zhao J, Libert T, Shadbolt N. Third party tracking in the mobile ecosystem. Accessed September 1, 2019. https://arxiv.org/pdf/1804.03603.pdf

12. Reyes I, Wijesekera P, Reardon J, et al "Won't somebody think of the children?" examining COPPA compliance at scale. *Proc Priv Enh Technol*. 2018; 2018:63-83. doi:10.1515/popets-2018-0021

13. Lee H, Wong SF, Oh J, Chang Y. Information privacy concerns and demographic characteristics: Data from a Korean media panel survey. *Gov Inf Q*. 2019;36:294-303. doi:10.1016/j.giq.2019.01.002

14. Park YJ. Digital literacy and privacy behavior online. *Commun Res.* 2013;40:215-236. doi:10.1177/0093650211418338

15. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381. doi:10.1016/j.jbi.2008.08.010

16. Radesky JS, Weeks HM, Ball R, et al. Young children's use of smartphones and tablets. *Pediatrics*. 2020;146(1):e20193518. doi:10.1542/peds.2019-3518

17. Livingstone S, Third A. *Children and Young People's Rights in the Digital Age: An Emerging Agenda.* Sage Publications; 2017.

18. Radesky J, Chassiakos YL, Ameenuddin N, Navsaria D. Digital advertising to children. *Pediatrics*. 2020;146(1):e20201681. doi:10.1542/peds.2020-1681

19. Kang R, Dabbish L, Fruchter N, Kiesler S. "My data just goes everywhere:" user mental models of the internet and implications for privacy and security. Paper presented at: Eleventh Symposium On Usable Privacy and Security; July 22-24, 2015; Ottawa, Canada. Accessed September 1, 2019. https://www.usenix.org/conference/soups2015/proceedings/presentation/kang

20. Schaub F, Breaux TD, Sadeh N. Crowdsourcing privacy policy analysis: potential, challenges and best practices. *it-Inf Technol.* 2016;58:229-236. doi:10.1515/itit-2016-0009

21. Apple. Privacy policy. Updated December 31, 2019. Accessed February 1, 2020. https://www.apple.com/legal/privacy/en-ww/