

RESEARCH ARTICLE

Recall accuracy of weekly automated surveys of health care utilization and infectious disease symptoms among infants over the first year of life

Catherine Ley^{1*}, Lauren Willis¹, Maria de la Luz Sanchez¹, Julie Parsonnet^{1,2}

1 Division of Infectious Diseases and Geographic Medicine, Department of Medicine, Stanford School of Medicine, Stanford, CA, United States of America, **2** Division of Epidemiology, Department of Health Research and Policy, Stanford School of Medicine, Stanford, CA, United States of America

* cley@stanford.edu



OPEN ACCESS

Citation: Ley C, Willis L, de la Luz Sanchez M, Parsonnet J (2019) Recall accuracy of weekly automated surveys of health care utilization and infectious disease symptoms among infants over the first year of life. *PLoS ONE* 14(12): e0226623. <https://doi.org/10.1371/journal.pone.0226623>

Editor: Florian Fischer, Bielefeld University, GERMANY

Received: May 6, 2019

Accepted: November 30, 2019

Published: December 17, 2019

Copyright: © 2019 Ley et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the National Institutes of Health, 5R01HD063142 and 5R21ES023371 (www.nih.gov) to JP and the Spectrum Pilot Seed Grant for Population Health Sciences, Stanford University (<http://med.stanford.edu/spectrum.html>) to JP. The funders had no role in study design, data collection and analysis,

Abstract

Automated surveys, by interactive voice response (IVR) or email, are increasingly used for clinical research. Although convenient and inexpensive, they have uncertain validity. We sought to assess the accuracy of longitudinally-collected automated survey responses compared to medical records. Using data collected from a well-characterized, prospective birth cohort over the first year of life, we examined concordance between guardians' reports of their infants' health care visits ascertained by weekly automated survey (IVR or email) and those identified by medical chart review. Among 180 survey-visit pairs, concordance was 51%, with no change as number of visits per baby increased. Accuracy of recall was higher by email compared to IVR (61 vs. 43%; adjusted OR = 2.5 95% CI: 1.3–4.8), did not vary by health care encounter type (hospitalization: 50%, ER: 64%, urgent care: 44%, primary care: 52%; $p = 0.75$), but was higher for fever (77%, adjusted OR = 5.1 95%CI: 1.5–17.7) and respiratory illness (58%, adjusted OR = 2.9 95%CI: 1.5–5.8) than for other diagnoses. For the 75 mothers in these encounters, 69% recalled at least one visit; among 41 mothers with two or more visits, 85% recalled at least one visit. Predictors of accurate reporting by mothers after adjusting for illness in the baby included increased age and increased years of education (age per year, $\beta = 0.05$, $p = 0.03$; education per year, $\beta = 0.08$, $p = 0.04$). Additional strategies beyond use of automated surveys are needed to ascertain accurate health care utilization in longitudinal cohort studies, particularly in healthy populations with little motivation for accurate reporting.

Introduction

Telephone surveys conducted by interviewers are ubiquitous, collecting information across a vast spectrum of domains—voter turnout, labor statistics, market research, healthcare utilization, and many others. Their pervasiveness is due to their presumed high level of accuracy, at

decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

least for objective phenomena [1]. Although interviewer-administered surveys are considered a gold standard for data collection, they are highly resource-intensive. Automated surveys via mobile or web-based technologies—for example, by interactive voice response (IVR, also known as robocalls) or email—have the distinct advantage of being convenient, efficient, and comparatively inexpensive. Automated surveys in medical care are generally well accepted: patient satisfaction, for example, was shown to be increased in patients who received a follow-up call after an ambulatory visit, with no difference between responses to a survey administered by a human caller or by IVR [2]. In some cases, IVR may in fact produce improved response rates compared to interviewer-administered surveys, particularly with respect to sensitive issues (e.g., depression in adolescents [3] or daily alcohol consumption [4, 5]).

Prospective cohort studies depend on the validity of self-report to make accurate and reliable conclusions. Validation of IVR surveys compared to self-administered paper questionnaires or interview by a clinician has been performed for a range of conditions including, among others: weight and height in young adults [6], neuropsychological testing [7]; teen addiction [8]; inflammatory bowel disease [9]; and depression [10–12]. Although several studies have evaluated the accuracy of in-person interviews or self-administered questionnaires to medical records with respect to health care utilization [13–17], pregnancy/birth history [18–22] and medical conditions [23–27], automated survey results have not been similarly compared to medical records either at a single time point or across time [28]. We sought to assess the accuracy of longitudinally-collected automated surveys in comparison to medical records over the first year of life in a multiethnic birth cohort, Stanford's Outcomes Research in Kids (STORK).

Materials and methods

Recruitment, enrollment and follow-up

Detailed methods for the STORK cohort have been described previously [29]. In brief, pregnant women were recruited through attendance at public obstetric clinics or by self-referral. An interested participant was eligible if fluent in English or Spanish and had a mobile phone. Written informed consent was administered at a baseline home visit convenient to the participant; written parental/guardian informed consent for the baby's participation was obtained once the baby was born. Information collected at the baseline household visit included demographic and household characteristics. Mothers were also asked to provide permission to review their medical records for infectious illnesses during pregnancy and for those in their infant after birth.

A short (less than three minute-long) automated survey (S1 Table) [29] was administered every week after delivery up until age three years regarding the baby's health. If the mother reported illness in the baby during the prior week, the survey asked about visits to a health care provider (HCP), infectious disease symptoms with their duration, and antibiotic use. If the baby was well, the survey instead asked an identical number of questions as the illness survey but about breastfeeding status, introduction to new foods, and amount of sleep. At the beginning of our study, in 2011, IVR surveys were used; shortly thereafter, however, some participants requested an email version for its greater convenience. Mothers could select or switch between any method (IVR, email or person-to-person) as they preferred, either by contacting the study office or by request at any 4-monthly household visit. Automated telephone calls were placed every 2 hours both Tuesday from 9:30 to 17:30 and Wednesday from 11:30 to 19:30 unless answered; emailed surveys were sent every Tuesday, Wednesday and Thursday at 7:30 unless answered. If no survey response was received after three days, study staff attempted to contact the participant that same week to administer the survey over the telephone.

Compensation for participation included free wash products, a \$10 gift card per participant per household visit and a \$25 gift card for each 16 weekly surveys completed. The study was approved by Stanford University's Administrative Panel on Human Subjects in Medical Research and the Santa Clara Valley Medical Center Institutional Review Board (ClinicalTrials.gov identifier: NCT01442701) and supported by the National Institutes of Health [5R01HD063142 and 5R21ES023371].

Chart review from any available medical records was performed to assess diagnoses and any associated prescriptions. Weekly automated surveys were administered using Precision Polling (Palo Alto, CA) if by telephone or Qualtrics (Provo, UT) if by email. Medical record information was abstracted into Medrio (San Francisco, CA) from either electronic medical records or from paper records obtained from HCPs. All data were managed in REDCap [30] hosted at Stanford University.

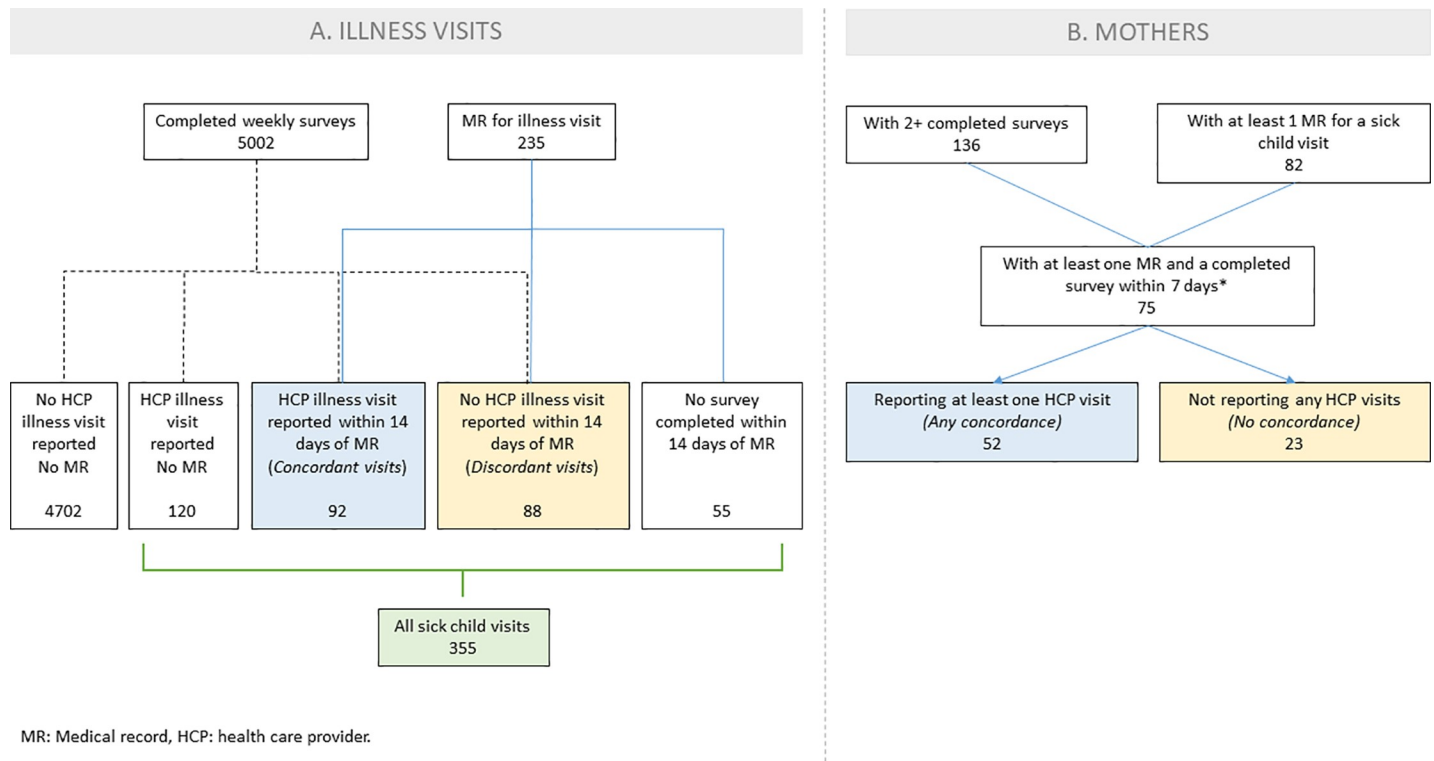
Statistical methods

Key outcomes. Because we obtained medical records only from mothers' and babies' primary health care facility, and because participants were free to visit unrelated emergency rooms, urgent care and other providers, mothers' reports of health care visits for illness could not always be confirmed. For this study, then, we chose to evaluate whether recorded illness visits at the primary institution were corroborated by the mother's survey responses. Recorded illness visits were defined as HCP visits for an illness for which we had medical records ("illness visits"); well-child check-ups or scheduled visits were excluded. We identified all IVR or email surveys that had been completed within 14 days after the recorded HCP visit. If two records were available during this time, only the one most proximate to the illness was reviewed. A record—survey pair was defined as concordant when the first survey response within 14 days of the HCP visit reported that visit. If the first survey within 14 days did not report the visit, the result was considered discordant. If no survey response was received, the visit was excluded from further analysis (Fig 1A). In addition to illness visits, we also chose to evaluate mothers. Mothers who reported at least one HCP illness visit within 14 days after the visit were classified as concordant, while those who did not report any illness visits despite the presence of at least one medical record for illness were classified as discordant (Fig 1B).

Secondary outcomes included the concordance of either a diagnosis of fever or a temperature measurement of 100°F or greater in the medical record and the report of fever in the survey, and the increasing number of concordant visits in mothers across time.

Key exposure and co-variables. The primary independent variables were maternal demographic and household factors, including maternal age, years of education and race/ethnicity (White, non-Hispanic; Asian, Hispanic or non-Hispanic; Black, Hispanic or non-Hispanic; White or Other, Hispanic). Co-variables related to study logistics included: the number of days between the first and the last survey completed (survey-days), the number of surveys completed overall, the survey completion rate (surveys completed / weeks in study), the number of surveys where illness was reported (sick-surveys), the proportion of all completed surveys where illness was reported, the total number of medical records and the total number of HCP illness visits (either recorded or reported, or both).

Statistics. We tested differences between concordant and discordant visits or between mothers with and without at least one concordant record-survey pair using Wilcoxon rank-sum tests or t-tests for continuous variables and chi-square, Mantel-Haenszel chi-square or Fisher's exact tests for categorical variables. We assessed the correlation between variables using the Spearman correlation coefficient. We used multivariable logistic regression modeling



MR: Medical record, HCP: health care provider.

Fig 1. Visit and participant disposition. The analysis set included only visits for illness diagnosed by a HCP and documented in medical records (well-baby check-ups or scheduled visits were not considered); these 235 visits corresponded to a total of 75 babies. Blue boxes indicate: in A, the number of illness visits reported on the available surveys [concordant visits]; in B, the number of mothers reporting at least one illness visit [concordant mothers]. Yellow boxes indicate: in A, the number of illness visits not reported on the available surveys [discordant visits]; in B, the number of mothers who did not report any HCP illness visits [discordant mothers]. The green box indicates the total number of visits for illness either reported (concordant, discordant or neither) or documented (concordant, discordant or neither). HCP: health care provider; MR: medical record.

<https://doi.org/10.1371/journal.pone.0226623.g001>

to explore predictors of concordance among visits and multivariable regression modeling to examine predictors of increasing numbers of concordant visits among mothers.

All analyses were performed in SAS Version 9.4 (Cary, NC).

Results

A total of 301 HCP illness visits were reported among the 5002 weekly surveys completed over the first year of life for the 136 babies born into the STORK cohort (median number of HCP visits for illness reported: 2; Q1-Q3: 0–3, range 0–13; median number of surveys answered: 42.5, Q1-Q3: 29–46, range: 2–51, with no difference in use of IVR for those who withdrew from the cohort compared to those who did not) (Fig 1, S2 Table). Of these reported HCP illness visits, 120 (40%) could not be confirmed by a medical record (median number of such survey reports: 1; Q1-Q3: 1–3, range: 1–10). Separately, a total of 235 illness visits were reported in the medical records of 82 babies (median number of records: 1; Q1-Q3: 1–4, range: 1–12). Of these illness visits in the medical records, 55 occurred during a two week period that was not covered by a survey (median number of such records: 1; Q1-Q3: 1–2, range: 1–8). The remaining 180 medical records, from 75 individual babies, pertained to visits that occurred during a two week period when a survey had been completed, and constituted our analysis sample.

Table 1. Concordance/discordance by diagnosis and location.

	Overall (% of total)	Concordant N (% of row)	P*
	180 (100)	92 (51.1)	
Data collection method			
IVR	99 (55.0)	43 (43.4)	0.07
Email	71 (39.4)	43 (60.6)	
Person-to-person call	10 (5.6)	6 (60.0)	
Location			
Hospitalizations	4 (2.2)	2 (50.0)	0.75
Emergency room	14 (7.8)	9 (64.3)	
Primary care	119 (66.1)	62 (52.1)	
Urgent care	39 (21.7)	17 (43.6)	
Other ⁽¹⁾	4 (2.2)	2 (50)	
Diagnoses			
Respiratory infections ⁽²⁾	113 (62.8)	65 (57.5)	0.03
Fever ⁽³⁾	17 (9.4)	13 (76.5)	0.03
Gastroenteritis	16 (8.9)	8 (50.0)	0.93
Conjunctivitis	10 (5.6)	2 (20.0)	0.05
Allergy ⁽⁴⁾	27 (15.0)	13 (48.2)	0.74
Non-allergic rash	21 (11.7)	10 (47.6)	0.73
Other ⁽⁵⁾	32 (17.8)	14 (43.8)	0.36

* Fisher's Exact test

⁽¹⁾ Other location: pediatric gastroenterology, dermatology clinic, telephone call to HCP

⁽²⁾ Respiratory infections: bronchitis; bronchiolitis; cold; congested/congestion; cough; croup; early viral infection; lower respiratory tract infection; nasopharyngitis; other viral syndrome; otitis media; pharyngeal erythema; possible pertussis; pneumonia; nasal rhinitis; sinusitis; stuffy nose; URI.

⁽³⁾ Mothers reported the occurrence of a fever for 35 of 46 (76%) visits where the chart indicated either a temperature measurement $\geq 100^\circ\text{F}$ and/or a diagnosis of fever. Two of these mothers reported the fever but not the HCP visit.

⁽⁴⁾ Allergy: acute hypersensitivity reaction; allergic rhinitis; amoxicillin rash; atopic dermatitis; eczema; food allergy; neonatal acne; non-specific skin eruption; wheezing/reactive airways disease.

⁽⁵⁾ Other diagnoses: anemia, aphtaes ulcers, life-threatening event, aspiration/vitamin E supplementation; blepharitis; bloating; blood stool; breath holding; congenital lacrimal stenosis; constipation; corneal abrasion; cradle cap; diarrhea; dacryocystitis; dacryostenosis; drainage of abscess left buttock; elevated lead; emesis; feeding problems; paronychia; fussiness; tracheomalacia; lymphadenitis; lymphadenopathy; penile irritation; pruritus; post feed emesis; rash; seborrhea; teething; urticaria.

<https://doi.org/10.1371/journal.pone.0226623.t001>

Visits

Mothers in this analysis selected a single method for responding to surveys, with only one woman switching from IVR to email responses over this first year of life. Just over half of surveys were collected via IVR (55%) (Table 1). Accuracy in reporting HCP visits for illness was higher for visits reported by email survey and by person-to-person survey compared to IVR (61% and 60% vs. 43%, $p = 0.07$). Mothers with least some college, compared to those with a high school education or less, and those aged greater than the median (31 years) compared to those aged less than the median, were more likely to select email surveys, have incurred more visits and have reported them (Fig 2); education was highly correlated with age ($r = 0.43$, $p < 0.001$). The total number of visits incurred did not affect the concordance rate ($p = 0.24$) (Table 2).

The majority of visits were to primary or urgent care (66% and 22%, respectively) (Table 1). Concordance by type of health care encounter did not vary across location type, with recall of a hospitalization or an emergency room visit equivalent to that for visits to a primary HCP and

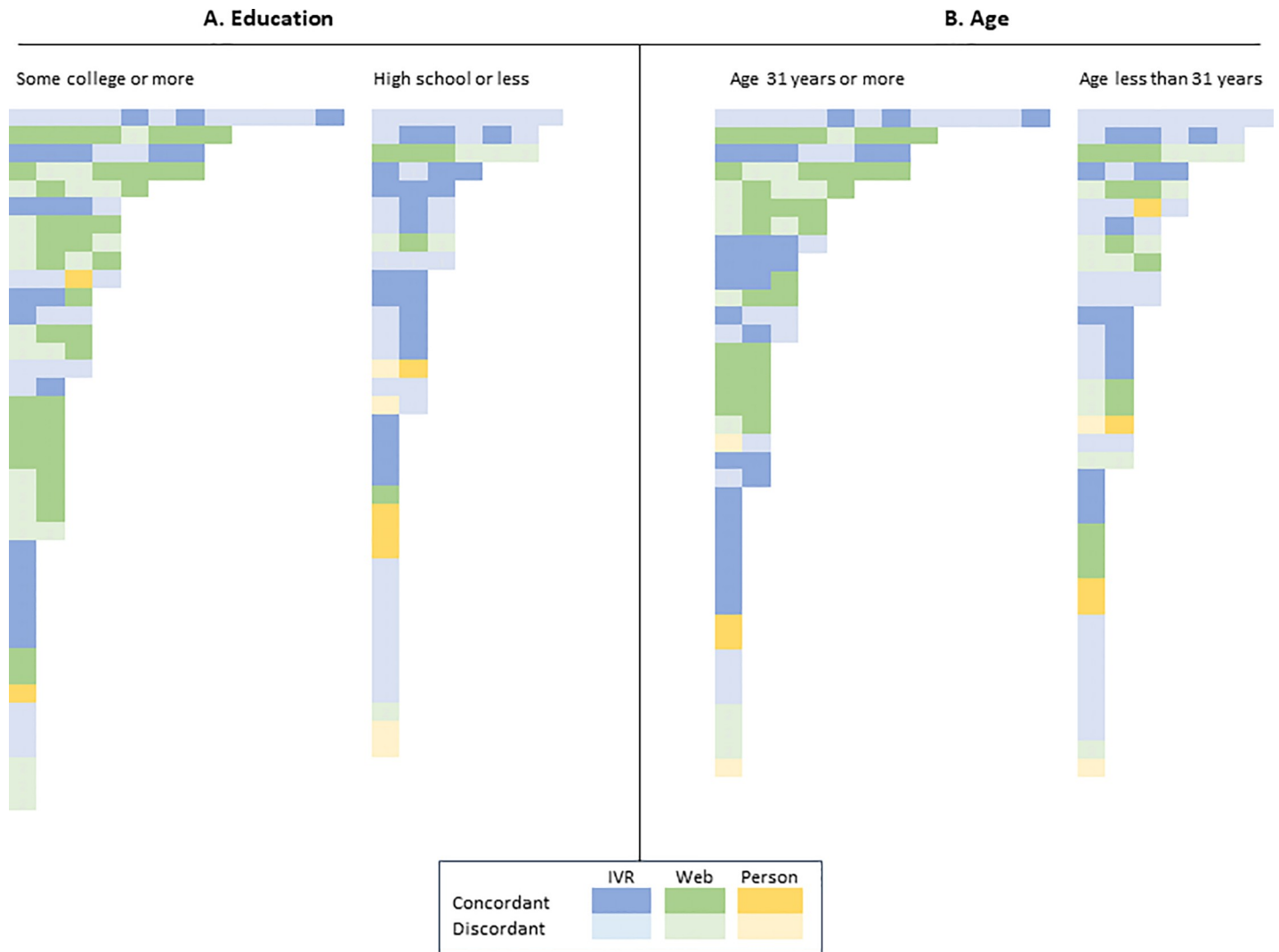


Fig 2. Distribution of visits per mother, stratified by (A) level of maternal education or (B) age. Education: any college vs. high school or less; age: less than vs. equal to or greater than the median (31 years). Each square represents one HCP illness visit. Each row within a column represents a mother. Within each column, mothers are sorted by total number of visits (highest to lowest). Visits within each mother are sorted by time, with the earliest visit on the left-hand side of the column. Visits are classified both by type of survey used after the visit (IVR: blue; email: green; person: yellow) and by concordance (concordant: dark; discordant: light).

<https://doi.org/10.1371/journal.pone.0226623.g002>

other sites, although hospitalizations were few ($N = 4$) ($p = 0.75$). Time since birth that the visit occurred was not associated with concordance ($p = 0.51$).

Of the 180 individual illness visits, 62.8% of all diagnoses (up to three recorded per visit) were related to respiratory infection (e.g., pneumonia, bronchiolitis, otitis media, upper respiratory infection [URI]), with allergic symptoms/diagnoses being the next most common (15.0%) (Table 1). Concordant visits were more likely than discordant visits to include a diagnosis of fever (76.5%, $p = 0.03$) or one related to respiratory infection (57.5%, $p = 0.03$) and less likely to include a primary diagnosis of conjunctivitis (20.0, $p = 0.05$).

Of the 17 diagnoses of fever identified in the medical report, 13 (76%) were subsequently reported by the mother on the weekly survey. Of the 46 instances of either a diagnosis of fever and / or a measured temperature equal to or greater than 100°F as recorded in the medical chart (28 instances had both, five had the diagnosis only), 35 (76.1%) were subsequently

Table 2. Agreement across numbers of survey-record pairs.

	Total	Concordant	%
Visits			
1	75 (42.7%)	32	0.43
2–3	65 (36.0%)	42	0.65
4–5	22 (12.2%)	9	0.41
6–12	18 (10%)	9	0.50
Overall	180	92	0.51*
Mothers with:			
1 visit	34 (45.3%)	17	0.50
2–3 visits	26 (34.7%)	21	0.81
4–5 visits	8 (10.7%)	8	1
6–12 visits	7 (9.3%)	6	0.86
Overall	75	52	0.69**

Test for linear trend

*p = 0.24

**p = 0.02

<https://doi.org/10.1371/journal.pone.0226623.t002>

reported by the mother (two of these occurrences of fever were reported even though the HCP visit was not).

Logistic regression modeling identified both illness type (fever and diagnoses related to respiratory infections) and survey method as independent predictors of concordance among visits (respiratory infections compared to none: OR = 2.9 95%CI: 1.5–5.8; fever compared to none: OR = 5.1 95%CI: 1.5–17.7; email compared to IVR: OR = 2.5 95% CI: 1.3–4.8).

Mothers

Based on demographic and household characteristics, the 75 babies and their mothers were similar to the entire STORK sample [29]. Mothers were on average approximately 30 years of age, primarily Hispanic (60.0%), with just over a third having their first child (Table 3). A total of 37% were born outside of the US, 35% spoke Spanish as their preferred language and approximately half had at most a high school education; their homes tended to be crowded, including on average four people of whom at least one was under the age of 18 years.

Of these 75 mothers, 52 (69%) had at least one concordant visit (median number of visits: 1, Q1-Q3:0–2, range: 0–7) (Table 2). Mothers with at least one concordant visit were slightly older than non-concordant mothers and had more years of education; they were also more likely to be enrolled longer in the cohort, answer more surveys and report more sick visits (Table 3). While white non-Hispanic mothers had a high concordance rate compared to other race/ethnic groups (90.9% vs. 60–67%), this distribution was not statistically significant (p = 0.39). Regression modeling identified both maternal age and maternal education as independent predictors of increasing numbers of concordant visits after adjustment for a measure of illness in the baby (total number of illness visits either reported, recorded or both) (age per year, $\beta = 0.05$, p = 0.03; education per year, $\beta = 0.08$, p = 0.04).

Discussion

Our goal in this study was to determine the agreement between any maternal report of an illness visit via automated weekly survey and the actual medical record from the HCP for infants followed over their first year of life. We found that overall only 51% of visits were accurately

Table 3. Demographic and household characteristics and study logistics, for mothers overall and with and without at least one concordant visit/report pair. Numbers presented are N (%) unless otherwise indicated.

		Overall	Concordant at least once		
		[% of total]	[% of row]		p ⁽¹⁾
		(N = 75)	Yes (N = 52)	No (N = 23)	
Demographic and household characteristics					
Maternal age (years)	Median (Q1-Q3)	31 (26–34)	31.5 (28.5–34)	28 (23–33)	0.05
Age category (years)	N (%)				0.17 ⁽²⁾
<26		17 (22.7)	10 (58.8)	7 (41.2)	
26–30		20 (26.7)	13 (65.0)	7 (35.0)	
31–33		16 (21.3)	12 (75.0)	4 (25.0)	
34+		22 (29.3)	17 (77.3)	5 (22.7)	
Race/ethnicity	N (%)				0.39 ⁽³⁾
W, NH		11 (14.7)	10 (90.9)	1 (9.1)	
A, H/NH		14 (18.7)	9 (64.3)	5 (35.7)	
B, H/NH		5 (6.7)	3 (60.0)	2 (40.0)	
W, H or O, H		45 (60.0)	30 (66.7)	15 (33.3)	
First born child	N (%)				0.61
No		49 (65.3)	33 (67.4)	16 (32.7)	
Yes		26 (34.7)	19 (73.1)	7 (26.9)	
Born in the US	N (%)				0.76
No		47 (62.7)	32 (68.1)	15 (31.9)	
Yes		28 (37.3)	20 (71.4)	8 (28.6)	
Preferred language	N (%)				0.99
English		49 (65.0)	34 (69.4)	15 (30.6)	
Spanish		26 (34.7)	18 (69.2)	8 (30.8)	
Years of education	Median (Q1-Q3)	13.0 (11–16)	14.0 (12–17.5)	12.0 (11–14)	0.03
Education category	N (%)				0.02 ⁽²⁾
Less than high school		22 (29.3)	12 (54.6)	10 (45.5)	
High school		14 (18.7)	9 (64.3)	5 (35.7)	
College (any)		22 (29.3)	16 (72.7)	6 (27.3)	
Post graduate		17 (22.7)	15 (88.2)	2 (11.8)	
Minors in household	Median (Q1-Q3)	1 (0–2)	1 (0–2)	1 (0–2)	0.98
Adults in household	Median (Q1-Q3)	3 (2–4)	3 (2–4)	2 (2–4)	0.18
Crowding	Median (Q1-Q3)	1.1 (0.8–1.7)	1.2 (0.8–1.7)	1.0 (0.8–1.5)	0.94
Baby sex	N (%)				0.15
Female		33 (44.0)	20 (60.6)	13 (39.4)	
Male		42 (56.0)	32 (76.2)	10 (23.8)	
Study logistics					
Time between first and last surveys (days)	Median (Q1-Q3)	336 (321–344)	338 (325–344)	323 (280–343)	0.03
Surveys answered	Median (Q1-Q3)	43 (34–46)	44 (41.5–47.5)	41 (22–43)	<0.001
Completion rate (surveys/weeks)	Median (Q1-Q3)	0.91 (0.83–0.96)	0.92 (0.88–0.97)	0.84 (0.76–0.93)	0.01
Sick visit surveys	Median (Q1-Q3)	3 (2–4)	3 (2–5)	2 (1–3)	0.005
Proportion of sick surveys (sick surveys / all surveys)	Median (Q1-Q3)	0.07 (0.04–0.12)	0.08 (0.04–0.13)	0.07 (0.04–0.09)	0.46
Medical records	Median (Q1-Q3)	2 (1–4)	3 (2–4)	2 (1–3)	0.04
All sick visits ⁽⁴⁾	Median (Q1-Q3)	3 (2–5)	4 (2–6)	2 (1–4)	0.03

⁽¹⁾ Compares mothers with concordant pairs to those without concordant pairs.

⁽²⁾ Mantel-Haenszel Chi-Square test

⁽³⁾ Fisher's Exact test

⁽⁴⁾ Any illness visit, including those reported by survey, by medical chart review or both.

<https://doi.org/10.1371/journal.pone.0226623.t003>

reported by the mother, with no difference across visit locations or with the total number of visits that occurred. Visits and reports for fever or respiratory infections (including otitis media) were more likely to be concordant than those for conjunctivitis, allergy or non-allergic rash, and reports and surveys by email or with an interviewer were more likely to be concordant than those by IVR. Among mothers, a total of 69% accurately reported that their sick baby had been seen at least once by a clinician; agreement increased to more than 80% among older mothers, among those with higher levels of education, and among those with multiple records from HCP visits.

During their first year of life, more than half of infants in our cohort had three or fewer illness-related HCP visits as per IVR report compared to two or fewer per medical record review. This rate of illness is within the range of other studies in healthy babies [31–34]. The discrepancy between reports and records could have been due to either incomplete medical record ascertainment or over-reporting by mothers; we did not have records for visits at sites not affiliated with the primary HCP. Accuracy in parental recall of the type and frequency of health care service visits has been shown to be quite variable, with recall of hospitalizations generally considerably better than that of emergency department or outpatient visits [35]. In our sample, recall of a visit after one week was accurate in only 50% of four hospitalizations recorded by automated survey in the first year of life; this proportion was not statistically different from that for visits to emergency, urgent or primary care.

Variability in recall is associated in part with both the time frame and the research question of interest [15, 16, 36]. Additionally, whether recall is improved with increased severity of the condition may be highly sample- and methodology- dependent. In our generally healthy cohort, sick child visits for fever or respiratory conditions—9% and 63% of all visits, respectively—were recalled slightly more accurately than for other conditions, on the order of 77% and 58%, respectively. A study of randomly selected Group Health Cooperative members in Washington state examined the ability of parents to recall injuries in their young children by telephone interview and showed that recall after two weeks ranged from 93% for major injuries to 71% for minor ones; recall one year later ranged from 56% to 19% for major and minor injuries, respectively [37]. On the other hand, in a highly motivated Dutch cohort at high risk of atopic conditions, agreement between the in-person maternal report at a doctor's interview every six months compared to the medical record was above 90% (range 77–100%) for most childhood illnesses [26].

Prior studies have suggested that recall of health-care related events can be affected by a wide range of demographic factors. We did not find that the number of children in the household affected agreement rates, suggesting that in this cohort, mothers of first-borns were not overly obsessed about reporting their child's health nor were mothers with many children less likely to report what happened to their youngest. Confirming other studies, we found that education level was strongly associated with accuracy, with mothers with less than a high school education having the lowest agreement rate, and those with post-graduate education the highest. Education level has long been associated with both baseline and longitudinal participation in research studies [38–40]. It is possible that, despite an extensive informed consent process, our cohort participants' understanding of the purpose of research and what was expected in terms of their participation (e.g., the time commitment needed for an accurate weekly report) was not fully appreciated by those with fewer years of schooling. In support of this, non-concordant mothers completed fewer surveys during their time enrolled compared to concordant mothers (84 vs. 92% of surveys). Finally, maternal age was a predictor of concordance. This finding may simply be due to participation bias from our study's location in, or affiliation with, an academic research environment, with potential participants perhaps more familiar with research requirements. Alternatively, increased age may be a marker for both increased education and higher income [41], shown previously to be a predictor of agreement [42, 43].

The scientific literature contains very little information on the accuracy of IVR-collected reports compared to medical record review. A pilot study of an IVR tool used in a hospital setting in Ghana showed agreement between childhood illness symptoms reported by the parents and those observed by physicians of 84, 82, 84 and 76% for fever, cough, diarrhea and vomiting, respectively [44]. These findings suggest that in appropriate settings, IVR surveys can be highly accurate. Most IVR interventions to date, however, appear to focus on improving patient follow-up/reminders and medication adherence [45], rather than data collection over time as in our cohort. Our weekly automated survey was short, focusing on illness and any HCP visits associated with illness within the prior seven days and took fewer than three minutes to complete if by IVR. Nevertheless, accurate response rates were far lower than we had anticipated, calling into question the use of IVR methods for longitudinal studies, particularly in healthy populations that are not highly motivated to give accurate responses.

Limitations to our analysis include the fact that the medical records for each baby were not comprehensive, so we could not evaluate standard agreement measures (e.g., kappa statistics). Families within the San Francisco Bay Area have access to a vast number of medical care services and it was not possible to retrieve all visit records. Accordingly, the medical records that we did have were considered our "gold standard," and we defined concordance in this study as the proportion of medical records where the mother reported a HCP visit within the prior week. It is possible that the request to report only sick-child visits on the automated survey created confusion with well-baby or other regularly scheduled visits, so the mother may have underreported these visits. We limited our analysis, however, to consider only those matched pairs that were available. Finally, a key underlying concern with recall over time is that mothers with sick children may be simply be unable or unwilling to report accurately even if by automated survey, possibly due to the short time frame of interest. It would be important to assess whether the accuracy of a bi-weekly or a monthly survey would be improved over one administered weekly, as these time frames may be preferred [28].

In summary, our study in a multiethnic birth cohort suggests that the use of weekly IVR or email surveys to report childhood illness does not reflect health care visits accurately. Additional effort is needed to generate accurate reporting systems, particularly in healthy populations that might have little motivation for accurate reporting.

Supporting information

S1 File. Data dictionary. Data dictionary for data files [S2 File](#) and [S3 File](#). (XLSX)

S2 File. Data per mother. Per mother data including number of concordant visits. (CSV)

S3 File. Data per visit. Per visit data including survey type, location of visit, diagnoses and concordance status comparing the survey report to the medical record. (CSV)

S1 Table. Questions included in the weekly survey, as administered by IVR, email or person-to-person. (DOCX)

S2 Table. Survey collection methods and demographic characteristics of mothers in the cohort overall, and in those who withdrew within the first year of the study compared to those who did not. (DOCX)

Acknowledgments

We thank Jon Krosnick, the Frederick O. Glover Professor in Humanities and Social Sciences and Professor of Communication, Political Science, and Psychology at Stanford University for initial insight on this topic. This analysis was funded in part by a Spectrum Pilot Seed Grant for Population Health Sciences, Stanford University.

Author Contributions

Conceptualization: Catherine Ley, Julie Parsonnet.

Data curation: Lauren Willis, Maria de la Luz Sanchez.

Formal analysis: Catherine Ley.

Funding acquisition: Julie Parsonnet.

Investigation: Catherine Ley, Julie Parsonnet.

Methodology: Catherine Ley.

Project administration: Julie Parsonnet.

Resources: Julie Parsonnet.

Validation: Catherine Ley, Lauren Willis.

Visualization: Catherine Ley.

Writing – original draft: Catherine Ley.

Writing – review & editing: Catherine Ley, Lauren Willis, Maria de la Luz Sanchez, Julie Parsonnet.

References

1. Pew Research Center: What Low Response Rates Mean for Telephone Surveys <http://www.pewresearch.org/methods/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/>
2. Houser SH, Ray MN, Maisiak R, Panjamapirom A, Willing J, Schiff GD, et al. Telephone follow-up in primary care: can interactive voice response calls work? *Stud Health Technol Inform*. 2013; 192:112–6. PMID: [23920526](https://pubmed.ncbi.nlm.nih.gov/23920526/)
3. Andersson C, Danielsson S, Silfverberg-Dymling G, Löndahl G, Johansson BA. Evaluation of Interactive Voice Response (IVR) and postal survey in follow-up of children and adolescents discharged from psychiatric outpatient treatment: a randomized controlled trial. *SpringerPlus*. 2014 February 8; 3:77. <https://doi.org/10.1186/2193-1801-3-77> PMID: [24567883](https://pubmed.ncbi.nlm.nih.gov/24567883/)
4. Perrine MW, Mundt JC, Searles JS, Lester LS. Validation of daily self-reported alcohol consumption using interactive voice response (IVR) technology. *J Stud Alcohol Drugs*. 1995 September; 56(5):487–90.
5. Searles JS, Helzer JE, Rose GL, Badger GJ. Concurrent and retrospective reports of alcohol consumption across 30, 90 and 366 days: interactive voice response compared with the timeline follow back. *J Stud Alcohol*. 2002 May; 63(3):352–62. <https://doi.org/10.15288/jsa.2002.63.352> PMID: [12086136](https://pubmed.ncbi.nlm.nih.gov/12086136/)
6. Nikolaou CK, Hankey CR, Lean MEJ. Accuracy of on-line self-reported weights and heights by young adults. *Eur J Public Health*. 2017 Oct 1; 27(5):898–903. <https://doi.org/10.1093/eurpub/ckx077> PMID: [28633350](https://pubmed.ncbi.nlm.nih.gov/28633350/)
7. Miller DI, Talbot V, Gagnon M, Messier C. Administration of Neuropsychological Tests Using Interactive Voice Response Technology in the Elderly: Validation and Limitations. *Front Neurol*. 2013; 4: 10 <https://doi.org/10.3389/fneur.2013.00010>
8. Brodey BB, Rosen CS, Winters KC, Brodey IS, Sheetz BM, Steinfeld RR, et al. Conversion and validation of the Teen-Addiction Severity Index (T-ASI) for Internet and automated-telephone self-report administration. *Psychology of Addictive Behaviors*. 2005 March; 19(1):54–61.
9. Lam MY, Lee H, Bright R, Korzenik JR, Sands BE. Validation of interactive voice response system administration of the Short Inflammatory Bowel Disease Questionnaire. *Inflamm Bowel Dis* 2009 Apr; 15(4):599–607. <https://doi.org/10.1002/ibd.20803> PMID: [19023897](https://pubmed.ncbi.nlm.nih.gov/19023897/)

10. Kunugi H, Koga N, Hashikura M, Noda T, Shimizu Y, Kobayashi T, et al. Validation of computer-administered clinical rating scale: Hamilton Depression Rating Scale assessment with Interactive Voice Response technology—Japanese version. *Psychiatry Clin Neurosci*. 2013 May; 6(4):253–8.
11. Moore HK, Hughes CW, Mundt JC, Rush AJ, Macleod L, Emslie GJ, et al. A pilot study of an electronic, adolescent version of the quick inventory of depressive symptomatology. *J Clin Psychiatry*. 2007 September; 68(9):1436–40. <https://doi.org/10.4088/jcp.v68n0917> PMID: 17915985
12. Turvey C, Sheeran T, Dindo L, Wakefield B, Klein D. Validity of the Patient Health Questionnaire, PHQ-9, administered through interactive-voice-response technology. *J Telemed Telecare*. 2012 September; 18(6):348–51. <https://doi.org/10.1258/jtt.2012.120220> PMID: 22933480
13. Brown JB, Adams ME. Patients as reliable reporters of medical care process: Recall of ambulatory encounter events. *Medical Care*. 1992 May; 30(5):400–411. <https://doi.org/10.1097/00005650-199205000-00003> PMID: 1583918
14. Brijnzeels MA, van der Wouden JC, Foets M, Prins A, van den Heuvel WJ. Validity and accuracy of interview and diary data on children's medical utilisation in the Netherlands. *J Epidemiol Community Health*. 1998 Jan; 52(1): 65–69. <https://doi.org/10.1136/jech.52.1.65> PMID: 9604044
15. D'Souza-Vazirani D, Minkovitz CS, Strobino DM. Validity of maternal report of acute health care use for children younger than 3 years. *Arch Pediatr Adolesc Med*. 2005 Feb; 159(2):167–72. <https://doi.org/10.1001/archpedi.159.2.167> PMID: 15699311
16. Liu G, Liao Z, Xu X, Liang Y, Xiong Y, Ni J. Accuracy of parent-reported measles-containing vaccination status of children with measles. *Public Health*. 2017 Mar; 144:92–95. <https://doi.org/10.1016/j.puhe.2016.12.013> PMID: 28274390
17. Nandy K, Menon U, Szalacha LA, Park H, Lee J, Lee EE. Self-Report Versus Medical Record for Mammography Screening Among Minority Women. *West J Nurs Res*. 2016 Dec; 38(12):1627–1638. <https://doi.org/10.1177/0193945916647059> PMID: 27138447
18. Bonner R, Bountziouka V, Stocks J, Harding S, Wade A, Griffiths C, et al. Birth data accessibility via primary care health records to classify health status in a multi-ethnic population of children: an observational study. *NPJ Prim Care Respir Med*. 2015 Jan 22; 25:14112. <https://doi.org/10.1038/npjpcrm.2014.112> PMID: 25612149
19. Hinkle SN, Rawal S, Zhu Y, Grewal J, Albert PS, Zhang C. Validation of self-reported diagnosis of gestational diabetes at 6-weeks postpartum. *Epidemiology*. 2017 Sep; 28(5):747–752. <https://doi.org/10.1097/EDE.0000000000000695> PMID: 28570385
20. Keenan K, Hipwell A, McAloon R, Hoffmann A, Mohanty A, Magee K. Concordance between maternal recall of birth complications and data from obstetrical records. *Early Hum Dev*. 2017 Feb; 105:11–15. <https://doi.org/10.1016/j.earlhumdev.2017.01.003> PMID: 28095344
21. Krakowiak P, Walker CK, Tancredi DJ, Hertz-Picciotto I. Maternal recall versus medical records of metabolic conditions from the prenatal period: a validation study. *Matern Child Health J*. 2015 September 19(9):1925–35. <https://doi.org/10.1007/s10995-015-1723-0> PMID: 25656730
22. Shenkin SD, Zhang MG, Der G, Mathur S, Mina TH, Reynolds RM. Validity of recalled v. recorded birth weight: a systematic review and meta-analysis. *J Dev Orig Health Dis*. 2017 Apr; 8(2):137–148. <https://doi.org/10.1017/S2040174416000581> PMID: 27776565
23. D'Aloisio AA, Nichols HB, Hodgson ME, Deming-Halverson SL, Sandler DP. Validity of self-reported breast cancer characteristics in a nationwide cohort of women with a family history of breast cancer. *BMC Cancer*. 2017 Oct 23; 17(1):692. <https://doi.org/10.1186/s12885-017-3686-6> PMID: 29058598
24. Camplain R, Kucharska-Newton A, Loehr L, Keyserling TC, Layton JB, Wruck L, et al. Accuracy of Self-Reported Heart Failure. The Atherosclerosis Risk in Communities (ARIC) Study. *J Card Fail*. 2017 Nov; 23(11):802–808. <https://doi.org/10.1016/j.cardfail.2017.09.002> PMID: 28893677
25. Cho S, Shin A, Song D, Park JK, Kim Y, Choi JY, et al. Validity of self-reported cancer history in the health examinees (HEXA) study: A comparison of self-report and cancer registry records. *Cancer Epidemiol*. 2017 Oct; 50(Pt A):16–21. <https://doi.org/10.1016/j.canep.2017.07.010> PMID: 28763723
26. Vissing NH, Jensen SM, Bisgaard H. Validity of information on atopic disease and other illness in young children reported by parents in a prospective birth cohort study. *BMC Med Res Methodol*. 2012 Oct 22; 12:160. <https://doi.org/10.1186/1471-2288-12-160> PMID: 23085330
27. Weiss A, Sommer G, Kuonen R, Scheinmann K, Grotzer M, Kompis M, et al. Validation of questionnaire-reported hearing with medical records: A report from the Swiss Childhood Cancer Survivor Study. *PLoS One*. 2017 Mar 23; 12(3):e0174479. <https://doi.org/10.1371/journal.pone.0174479> PMID: 28333999
28. Dreyer NA, Blackburn SC, Mt-Isa S, Richardson JL, Thomas S, Laursen M, et al. Direct-to-Patient Research: Piloting a New Approach to Understanding Drug Safety During Pregnancy. *JMIR Public Health Surveill*. 2015 Dec 22; 1(2):e22. <https://doi.org/10.2196/publichealth.4939> PMID: 27227140

29. Ley C, Sanchez M de la L, Mathur A, Yang S, Sundaram V, Parsonnet J. Stanford's Outcomes Research in Kids (STORK): a prospective study of healthy pregnant women and their babies in Northern California. *BMJ Open*. 2016 April 13; 6(4):e010810 <https://doi.org/10.1136/bmjopen-2015-010810> PMID: 27075843
30. 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*. 2008 September 30; 42.
31. Aglipay M, Birken CS, Parkin PC, Loeb MB, Thorpe K, Chen Y, et al. Effect of High-Dose vs Standard-Dose Wintertime Vitamin D Supplementation on Viral Upper Respiratory Tract Infections in Young Healthy Children. *JAMA*. 2017 Jul 18; 318(3):245–254. <https://doi.org/10.1001/jama.2017.8708> PMID: 28719693
32. Schuez-Havupalo L, Lahti E, Junttila N, Toivonen L, Aromaa M, Rautava P, et al. Parents' depression and loneliness during pregnancy and respiratory infections in the offspring: A prospective birth cohort study. *PLoS One*. 2018 Sep 7; 13(9):e0203650. <https://doi.org/10.1371/journal.pone.0203650> PMID: 30192872
33. Sarna M, Ware RS, Sloots TP, Nissen MD, Grimwood K, Lambert SB. The burden of community-managed acute respiratory infections in the first 2-years of life. *Pediatr Pulmonol*. 2016 Dec; 51(12):1336–1346. <https://doi.org/10.1002/ppul.23480> PMID: 27228308
34. Toivonen L, Karppinen S, Schuez-Havupalo L, Teros-Jaakkola T, Vuononvirta J, Mertsola J, et al. Burden of Recurrent Respiratory Tract Infections in Children: A Prospective Cohort Study. *Pediatr Infect Dis J*. 2016 Dec; 35(12):e362–e369. <https://doi.org/10.1097/INF.0000000000001304> PMID: 27455443
35. Dragojlovic N, Kim E, Elliott AM, CAUSES Study, Friedman JM, Lynd LD. Evaluating the use of parental reports to estimate health care resource utilization in children with suspected genetic disorders. *J Eval Clin Pract*. 2018 Apr; 24(2):416–422. <https://doi.org/10.1111/jep.12876> PMID: 29603523
36. Kjellsson G, Clarke P, Gerdtham UG. Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions. *J Health Econ*. 2014 May; 35:34–46. <https://doi.org/10.1016/j.jhealeco.2014.01.007> PMID: 24595066
37. Cummings P, Rivara FP, Thompson RS, Reid RJ. Ability of parents to recall the injuries of their young children. *Inj Prev*. 2005; 11:47–47
38. Cobb EM, Singer DC, Davis MM. Public interest in medical research participation: differences by volunteer status and study type. *Clin Transl Sci*. 2014 Apr; 7(2):145–9. <https://doi.org/10.1111/cts.12142> PMID: 24456538
39. Lissner L, Skoog I, Andersson K, Beckman N, Sundh V, Waern M, et al. Participation bias in longitudinal studies: experience from the population study of women in Gothenburg, Sweden. *Scand J Prim Health Care*, 2003 21:4, 242–247
40. Unger JM, Hershman DL, Albain KS, Moynihan CM, Petersen JA, Burg K, et al. Patient income level and cancer clinical trial participation. *J Clin Oncol*. 2013 Feb 10; 31(5):536–42. <https://doi.org/10.1200/JCO.2012.45.4553> PMID: 23295802
41. Galobardes B, Shaw M, Lawlor DA, Lynch JW. Indicators of socioeconomic position (part 1). *J Epidemiol Community Health* 2006 60(1): 7–12 <https://doi.org/10.1136/jech.2004.023531> PMID: 16361448
42. Burakevych N, McKinlay CJ, Alswelker JM, Harding JE, CHYLD Study team. Accuracy of caregivers' recall of hospital admissions: implications for research. *Acta Paediatr*. 2015 Nov; 104(11):1199–204. <https://doi.org/10.1111/apa.13208> PMID: 26355393
43. McPhee SJ, Nguyen TT, Shema SJ, Nguyen B, Somkin C, Vo P, et al. Validation of recall of breast and cervical cancer screening by women in an ethnically diverse population. *Prev Med*. 2002 Nov; 35(5):463–73. <https://doi.org/10.1006/pmed.2002.1096> PMID: 12431895
44. Franke KH, Krumkamp R, Mohammed A, Sarpong N, Owusu-Dabo E, Brinkel J, et al. A mobile phone based tool to identify symptoms of common childhood diseases in Ghana: development and evaluation of the integrated clinical algorithm in a cross-sectional study. *BMC Med Inform Decis Mak*. 2018 Mar 27; 18(1):23 <https://doi.org/10.1186/s12911-018-0600-3> PMID: 29580278
45. Tsoli S, Sutton S, Kassavou A. Interactive voice response interventions targeting behaviour change: a systematic literature review with meta-analysis and meta-regression. *BMJ Open*. 2018 Feb 24; 8(2):e018974. <https://doi.org/10.1136/bmjopen-2017-018974> PMID: 29478016