

Survey non-response in an internet-mediated, longitudinal autism research study

Luther G Kalb,¹ Cheryl Cohen,² Harold Lehmann,³ Paul Law^{4,5}

¹Center for Autism and Related Disorders, Kennedy Krieger Institute, Baltimore, Maryland, USA

²Department of Medical Informatics, Kennedy Krieger Institute, Baltimore, Maryland, USA

³Division of Health Science Informatics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

⁴Department of Medical Informatics, Kennedy Krieger Institute, Baltimore, Maryland, USA

⁵Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Correspondence to

Dr Paul Law, Department of Medical Informatics, Kennedy Krieger Institute, 3825 Greenspring Avenue, Painter Building, 1st Floor, Baltimore, MD 21211, USA; lawp@kennedykrieger.org

Received 19 January 2012

Accepted 24 March 2012

Published Online First

26 April 2012

ABSTRACT

Objective To evaluate non-response rates to follow-up online surveys using a prospective cohort of parents raising at least one child with an autism spectrum disorder. A secondary objective was to investigate predictors of non-response over time.

Materials and Methods Data were collected from a US-based online research database, the Interactive Autism Network (IAN). A total of 19 497 youths, aged 1.9–19 years (mean 9 years, SD 3.94), were included in the present study. Response to three follow-up surveys, solicited from parents after baseline enrollment, served as the outcome measures. Multivariate binary logistic regression models were then used to examine predictors of non-response.

Results 31 216 survey instances were examined, of which 8772 or 28.1% were partly or completely responded to. Results from the multivariate model found non-response of baseline surveys (OR 28.0), years since enrollment in the online protocol (OR 2.06), and numerous sociodemographic characteristics were associated with non-response to follow-up surveys (all $p < 0.05$).

Discussion Consistent with the current literature, response rates to online surveys were somewhat low. While many demographic characteristics were associated with non-response, time since registration and participation at baseline played the greatest role in predicting follow-up survey non-response.

Conclusion An important hazard to the generalizability of findings from research is non-response bias; however, little is known about this problem in longitudinal internet-mediated research (IMR). This study sheds new light on important predictors of longitudinal response rates that should be considered before launching a prospective IMR study.

response error as ‘a function both of the rate of non-response (the proportion not responding over the total eligible for the survey) and the differences between those who respond and those who do not on the variable of interest’ (p. 87). Several studies have shown non-response rates to be high in survey-based IMR,^{12–14} although there is great variability in these rates due to differences in sampling methodology. For instance, response rates from 6%¹⁵ to 75%¹⁶ have been reported for email-based surveys, while rates as low as 0.26% have been identified for surveys solicited of website visitors.¹⁷ In a meta-analysis of web-response rates, Cook *et al*¹⁸ found a mean response rate of approximately 40%. When comparing traditional versus IMR response rates, a meta-analysis by Manfreda and colleagues¹⁹ showed that the response rates in web-based surveys were 11% lower than other survey modalities. However, Greenlaw and Brown-Welty,²⁰ using an experimental design, showed that surveys administered by mixed modes (both web-based and conventional mail-based) to be the most effective, with a response rate of 42% for paper-based surveys, 52% for web-based surveys, and 60% for mixed-mode surveys. As a sample becomes more biased (and less generalizable) as the non-response rate increases, this bias should be addressed when disseminating findings from IMR.

Several cross-sectional studies have examined factors associated with survey response. For survey-based IMR, participant factors associated with response rates include age, socioeconomic status, rurality or urbanicity, health status or disease severity, ethnicity/race, gender, and self-efficacy.^{21–24} Study-related factors are also important. These include site aesthetics, confidentiality, appearance of legitimacy of research and institution, personalized contacts, multiple contacts, and precontact (ie, when researchers contact potential subjects before the survey is administered).^{18 21 25–27} For conventional mail, salience of the survey’s content, whether the survey was sent by recorded delivery versus standard delivery, and incentives have shown importance.²⁸

Although this body of research is growing, little is known about response rates over time or its predictors in longitudinal IMR. To our knowledge, only two internet-mediated survey-based studies have examined retention rates at more than one time point. Using a pre-post design, Sax *et al*²⁹ found a response rate of 17% for web surveys without a response incentive and 19.8% for those with an incentive in a large cohort of freshman across several US-based universities. Interestingly, there was little difference between internet-

As familiarity with and use of the internet increases, even across the digital divide,^{1 2} health researchers have found the online environment to be a viable mechanism for data collection within the context of research.^{3 4} Online research, sometimes referred to as internet-mediated research (IMR), is particularly attractive because of its low cost and superior sample representativeness when compared with traditional university or center-based research.^{5–7} As with any research design, IMR is not without its limitations and biases. Despite the well-known and well-documented concerns surrounding IMR,^{8–10} there is a paucity of research that addresses threats to validity facing survey-based IMR when compared with traditional, non-IMR research methodologies.

One of the key challenges confronting IMR is non-response bias or error. Couper¹¹ defines non-

mediated and mail-based response rates (22%). As for predictors of response, these included female gender, non-hedonistic behavior (eg, abstinence from drinking and/or smoking), increased socioeconomic status and SAT scores, and personality characteristics (eg, English/fine-arts majors, leadership qualities, and social activists). Furthermore, Khosropour and Sullivan³⁰ employed a 3-month pre-post design in an online study of sexual behavior among men who have sex with men. These authors found a 22% response rate and predictors of response included Caucasian ethnicity, financial incentive, and an active email account. While these studies shed some light on the topic, inconsistent results, data capture at only two time points, convenient sampling of undergraduate students and disparate populations limit the generalizability of these findings. Taken together, further research surrounding non-response error, with a particular emphasis on identifying unique factors associated with non-response to web-based surveys over time, is necessary as the utility of IMR continues to grow.

OBJECTIVES

The primary goal of the present study is to evaluate non-response rates to follow-up (ie, surveys launched after the registration and consent process) web-based surveys using a cohort of parents rearing at least one child with an autism spectrum disorder (ASD) engaged in a voluntary, longitudinal online research study. The second objective is to investigate what child and parent demographic factors are associated with survey non-response. The third aim is to explore survey exposure (ie, time until survey response since original launch or solicitation) and time since registration with the online research protocol as factors that may influence response rates. Fourth and finally, an important question to longitudinal IMR research will be examined. That is, to what effect does engagement, or lack thereof, in the online protocol at baseline (ie, never filled out any surveys at initial registration) have on future response rates?

MATERIALS AND METHODS

Setting

Data were collected from a US-based online research database, the Interactive Autism Network (IAN). This unique research mechanism is designed to foster collaboration between the autism community and investigators while assisting with overcoming traditional barriers to research. IAN, launched in April 2007, is now the largest online autism research effort. A total of 19 497 youths, aged 1.9–19 years (mean 9 years, SD 3.94), were included in the present study.

Designed as a longitudinal protocol, IAN engages families over the lifespan through two primary mechanisms: IAN Community and IAN Research. IAN Community (<http://www.IANcommunity.org>) is a website where the public learns and discusses autism and autism-related research. More specifically, the website provides articles by leaders in the field, discussion forums focused on recent research and the research process, and preliminary findings from IAN Research so that participants can see the value of their contributions.

While IAN Community engages the community in the research enterprise through an informational website and ongoing discussion, IAN Research gathers information from families. The resulting data are both used by the IAN Research team and shared throughout the research community after de-identification procedures take place for Health Insurance Portability and Accountability Act of 1996 (HIPAA) compliance. In addition, the data are used to help match IAN Research participants with

studies for which they qualify. To date, IAN has provided subject recruitment assistance for nearly 300 autism research projects.

IAN Research, which is located at <http://www.IANresearch.org>, currently collects four types of data: registration, baseline, IAN longitudinal treatment protocol, and survey. All individuals who have been diagnosed with an ASD along with certain family members are qualified to participate. For probands (ie, the affected child) and siblings who are under 18 years of age, a parent or legal guardian registers and consents his/her dependents and himself/herself. Based on that registration data, the IMR system assigns appropriate baseline surveys and a longitudinal protocol pertaining to proband, siblings, and parents. Surveys are administered as needed and research participation is ongoing. In addition, participants receive notifications to participate in third-party studies until they ask to withdraw from the IAN study. Note that there is also a protocol for adult probands, which was not used in the current analysis.

Measures

Primary outcome

The primary outcome variable for this study was survey non-response, with response being defined as partial or total completion of the survey instance. Given 95% of those who started the survey also completed it, non-response represents parents who did not open or start the survey instance.

Baseline surveys

After a family completes the registration and consent process, the IMR system assigns baseline surveys and initial longitudinal surveys for the child with ASD, parents or guardians, and unaffected siblings. The baseline surveys consist of questionnaires developed by IAN researchers and collaborators as well as standardized instruments. Surveys developed by IAN researchers and collaborators include the Sibling without ASD, Child with ASD, Mother Basic Medical History, Father Basic Medical History, and IAN longitudinal treatment protocol. The standardized instruments include the Social communication questionnaire³¹ and the Social Responsiveness Scale.³²

Outcome surveys

In addition to the baseline and longitudinal surveys, IAN Research periodically administers one-time surveys to fill strategic gaps in ASD knowledge. The Access to Healthcare Survey, Vaccination History, and Weight and Height Survey served this purpose and functioned as outcomes for this study. The Access to Healthcare Survey, which assessed access to healthcare services, was administered to the parent for the affected child. The Vaccination History Survey contained questions assessing the parent's vaccination beliefs, attitudes, and practices and was administered to the parent for the affected and unaffected children. The Weight and Height Survey gathered basic child height and weight parameters and was administered to the parent for the affected child and the unaffected siblings. Of note, each survey was open for different lengths of time (Access to Healthcare Survey, 154 days; Vaccination History Survey, 418 days; Weight and Height Survey, 32 days).

Demographic characteristics

Demographic items used in the current analysis were taken from the sibling without ASD, child with ASD registration, and initial registration surveys. These included gender, race/ethnicity, ASD diagnosis, number of affected and non-affected children, mother's education, family structure, child and parental age, and urbanicity. A mutually exclusive race variable using the following

two categories was created: white, and non-white (ie, African American, Hispanic, multiple, and other). Similarly, for mother’s highest level of education, the following three categories were created: up to high school graduate or equivalent, some college experience, and graduate-level education. Finally, a qualitative variable was created to reflect urbanicity using the 2006 National Center for Health Statistics rural–urban community area codes. The National Center for Health Statistics developed a six-level classification scheme based on the 2000 Census that ranged from the most urban category, consisting of large metropolitan central counties, to rural, non-metropolitan counties.³³

Data collection and analysis

All survey data entered by parents were collected and maintained using the IMR components of clinical research management system (MDLogix, Baltimore, Maryland, USA). Electronic consent and assent were obtained from all participants using methods approved by the Johns Hopkins Medicine Institutional Review Board. Stata V.11.0 was used to perform the data analysis on data extracted on 21 January 2011.

For the statistical analysis, bivariate analyses, using paired t and McNemar’s tests, for continuous and categorical data, were used first to examine differences between responders and non-responders. These tests, as well as the regression methodology that accounted for clustering of observations (see below for details), were chosen because the assumption of independence does not hold for the present study. That is, a participant could be represented in one (eg, non-response to all surveys) or both (eg, response to one survey and non-response to other surveys) outcome groups because more than one survey instance may have been solicited of the participant. Once significant ($p < 0.05$) comparisons from the bivariate analyses were identified, subsequent analyses using a multi-variate binary logistic regression model, which adjusted for repeated observations of each child (ie, clustering) using Stata’s clustered sandwich estimator,³⁴ was employed to develop odds ratio (OR). OR were used to examine the likelihood of an event (such as survey response) occurring in one group compared with the odds of it occurring in another (eg, graduate vs high school parental education). Backwards elimination was used to

Table 1 Demographic characteristics of respondents and non-respondents across three surveys

	Access to Healthcare		Vaccination History		Weight and Height	
	Respond		Respond		Respond	
	Y	N	Y	N	Y	N
N (%)	384 (26%)	2044 (84%)	4758 (25%)	14 348 (75%)	3630 (27%)	6052 (63%)
Mother’s age (mean, years)	41.1	39.1	39.6	38.8	38.8	38.4
Child’s age	9.6	9.4	9.1	9.2	8.4	8.7
Child’s gender						
Male	16%	84%	25%	75%	38%	62%
Female	16%	84%	25%	75%	37%	63%
Child’s race						
White	16%	84%	26%	74%	38%	62%
Non-white	13%	87%	20%	80%	33%	67%
Child’s ASD diagnosis						
No ASD	N/A	N/A	25%	75%	38%	62%
ASD			25%	75%	37%	63%
No of children						
1	14%	86%	22%	78%	33%	67%
2	17%	83%	28%	72%	40%	60%
3+	16%	84%	24%	76%	36%	64%
No of children with ASD						
1	17%	83%	26%	74%	38%	62%
2+	10%	90%	22%	78%	34%	66%
Mother’s education						
No college	15%	85%	25%	75%	35%	65%
Some college	16%	84%	27%	73%	39%	61%
Some graduate	27%	73%	33%	67%	40%	60%
Some graduate school						
Family structure						
2 Parents	17%	83%	27%	73%	40%	60%
1 Parent	13%	87%	20%	80%	24%	76%
Rurality						
1 Very rural	13%	87%	29%	71%	38%	62%
2	19%	81%	28%	72%	41%	59%
3	14%	86%	27%	73%	38%	62%
4	14%	86%	24%	76%	40%	60%
5	17%	83%	24%	76%	35%	65%
6 Dense metro	16%	84%	24%	76%	35%	65%
Average exposure until response						
Mean days (SD)	59 (54)		54 (92)		20 (61)	
Time since registration with IAN						
Years (SD)	1.78 (0.9)	2.05 (0.75)	1.03 (0.9)	1.3 (0.65)	0.47 (0.17)	0.45 (0.18)

ASD, autism spectrum disorder; IAN, Interactive Autism Network.

develop the final model with only those variables significant at $p < 0.05$ remaining.

For the third objective, a Kaplan–Meier procedure was used to estimate the proportion of parents who responded to a particular survey across time since exposure to the survey. We graphed the trend of response, as opposed to non-response, because this trend is more intuitive and easier to interpret. In this analysis, a participant was considered to be a censored observation if they had not responded to the survey by the time it closed.

There were very few missing data in the present study. In fact, fewer than 5% of cases were missing on any variable included in the analysis. Values for children whose parents either did not know or refused to answer a question were coded as missing. Two variables that were missing substantial data (>33%) were father’s education level and age. Due to this high proportion, these variables were omitted and mother’s information was used in the analysis.

RESULTS

Demographics

From three different surveys, 31 216 survey instances were examined, of which 8772 or 28.1% were respondents. Table 1 shows demographic information about the sample stratified by survey and response. No statistical tests are listed in table 1, as all of the surveys referred to in this table are aggregated and analyzed in table 2. The α value on the right side of table 2 presents the significance value for the difference between responders and non-responders across all survey instances; all of which, except for gender, were significant ($p < 0.05$).

Predictors of non-response to follow-up online survey instances

The multivariate model identified many factors associated with survey non-response. These included increasing child age, decreasing maternal age, more than one affected child with ASD, single parent households, lower maternal education, non-white families, increasing urbanicity, increasing duration since initial registration with IAN, and whether the participant filled out at least one baseline survey at registration (all $p < 0.05$). Table 3 displays the specific test statistics from the model.

Time to response for Access to Healthcare, Vaccination History, and Weight and Height Surveys

A Kaplan–Meier procedure was used to estimate the proportion of participants that responded to the survey since original solicitation or launch of questionnaire. Figure 1 shows there is a different effect of exposure or time to response by survey. For instance, more than 90% of participants who responded to the Weight and Height Survey had done so by 30 days. While, at 30 days, approximately one-third had responded to the Access to Healthcare Survey and approximately half had responded to the Vaccination History Survey. Finally, this graph shows the disparity in response rates, as well as the time to response, between the surveys.

Effect of time since registration in IAN on rate of non-response for baseline survey respondents and non-respondents

Eighteen per cent of participants did not respond to any surveys at baseline. The total non-response rate to follow-up surveys for these individuals was 98%. This is much higher when compared with the follow-up non-response rate of 66% for participants who responded to at least one baseline survey. Given this disparity, it is no surprise this variable was the strongest predictor in the multivariate model (see table 3).

Table 2 Demographic differences between respondents and non-respondents across all follow-up survey instances

	Respond		p Value
	Y	N	
N (%)	8772	22 444 (72%)	
Mother’s age (mean, years)	39.3	38.7	<0.001
Child’s age	9.09	8.8	<0.001
Child’s gender			0.07
Male	28%	72%	
Female	28%	72%	
Child’s race			<0.001
White	29%	71%	
Non-white	23%	77%	
Child’s ASD diagnosis			0.001
No ASD	34%	66%	
ASD	36%	64%	
No of children			<0.001
1	24%	76%	
2	32%	68%	
3+	27%	73%	
No of children with ASD			<0.001
1	29%	71%	
2+	24%	76%	
Mother’s education			<0.001
No college	27%	73%	
Some college	30%	70%	
Some graduate School	35%	65%	
Family structure			<0.001
2 Parents	31%	69%	
1 Parent	20%	80%	
Rurality			<0.001
1 Very rural	30%	70%	
2	31%	69%	
3	30%	70%	
4	29%	71%	
5	28%	72%	
6 Dense metro	27%	73%	
Average exposure until response			
Mean days (SD)	41 (81.5)		
Time since registration with IAN			<0.001
Years (SD)	0.84 (68)	1.14 (0.74)	

ASD, autism spectrum disorder; IAN, Interactive Autism Network.

Figure 2 shows the effect of time since registration to IAN on the non-response rate between these two populations. For individuals who responded to at least one baseline survey, there is an increasing monotonic relationship between time enrolled in IAN and non-response rate, although this changes over time. That is, the non-response rate dramatically increases in the first year of enrollment and decreases thereafter. For individuals who did not respond at baseline, non-response was extremely high and continued to stay that way throughout the course of their online experience. In the multivariate model, time since registration was the second strongest predictor of non-response (see table 3).

DISCUSSION

The present study examined non-response to web-based surveys in families of a child with an ASD engaged in an online, longitudinal research protocol. Consistent with the longitudinal IMR literature, overall non-response rates were somewhat high; warranting further exploration into the factors associated with this barrier to generalizable epidemiological research.

Table 3 Predictors of non-response to all follow-up survey instances

Variable	OR	SE	Z score	p Value	95% CI
Baseline survey response					
At least 1 baseline survey completed	Referent				
No baseline surveys complete	28.03	3.72	25.10	<0.001	21.6 to 36.4
Years since registration with IAN	2.06	0.04	35.25	<0.001	1.97 to 2.14
Marital status					
Married	Referent				
Single household	1.32	0.06	6.48	<0.001	1.21 to 1.43
Mother's education					
Some graduate school	Referent				
Some college	1.13	0.05	2.80	0.005	1.04 to 1.25
No college	1.22	0.08	3.17	0.002	1.08 to 1.39
Race					
White	Referent				
Non-white	1.12	0.06	2.17	0.03	1.01 to 1.24
No of children with ASD					
1 Child with ASD	Referent				
More than 1 child with ASD	1.13	0.05	2.92	0.003	1.04 to 1.22
Urbanicity					
0 Very rural	Referent				
1	0.96	0.09	-0.42	0.67	0.80 to 1.15
2	1.04	0.10	0.43	0.67	0.86 to 1.25
3	1.11	0.09	1.27	0.20	0.94 to 1.32
4	1.19	0.10	2.11	0.035	1.01 to 1.40
5 Dense metro	1.19	0.10	2.02	0.043	1.00 to 1.41
Child's age	1.02	0.005	4.81	<0.001	1.01 to 1.03
Mother's age	0.97	0.003	-9.13	<0.001	0.96 to 0.98

ASD, autism spectrum disorder; IAN, Interactive Autism Network.

Results from the multivariate analyses identified a number of demographic characteristics associated with non-response. The strongest predictors included single households, older children and younger mothers. Other significant, albeit much weaker, variables included increasing urbanicity, non-white families, rearing more than one child with an ASD, and lower maternal education, while child gender and the number of non-affected children were not significant in the model. Further research is needed to examine other, more in-depth child (eg, comorbidity, disease severity) and family (eg, parental stress, family quality of life, community support), factors that predict IMR engagement.

Although many demographic variables were identified as statistical predictors of non-response, the variable that contributed more variance, or understanding about non-response, than all other factors was baseline survey response behavior. Engagement at enrollment is important as IAN aims to retain

families for a variety of reasons, one of which is to collect longitudinal data. Even after controlling for demographics and time enrolled in the study, those who did not complete one survey at baseline were 28 times more likely not to respond, compared with those who do respond, to follow-up surveys. These data suggest that this pattern of non-response should be considered before launching a longitudinal IMR study, and specific attempts, perhaps through targeted interventions soon after enrollment, should be made to engage this population.

The second strongest determinant of survey non-response was time enrolled in the study. Results from the multivariate model show that the likelihood of non-response doubles for each year increase in study enrollment; even after adjusting for baseline non-responders and demographics. Figure 2 descriptively illustrates this finding by the steadily increasing percentage of non-response with the longer a family or parent was enrolled in IAN.

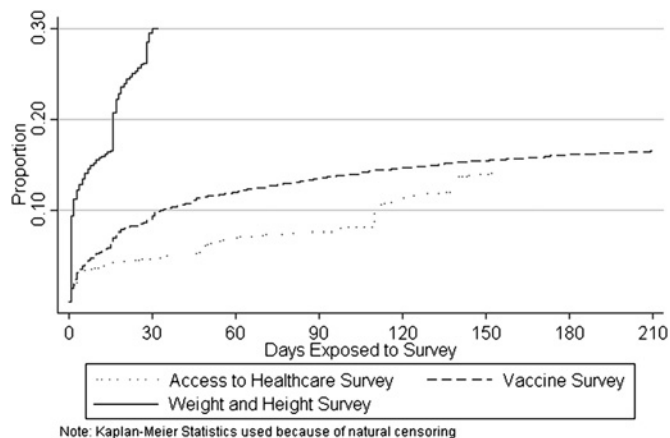


Figure 1 Time to response for all follow-up surveys.

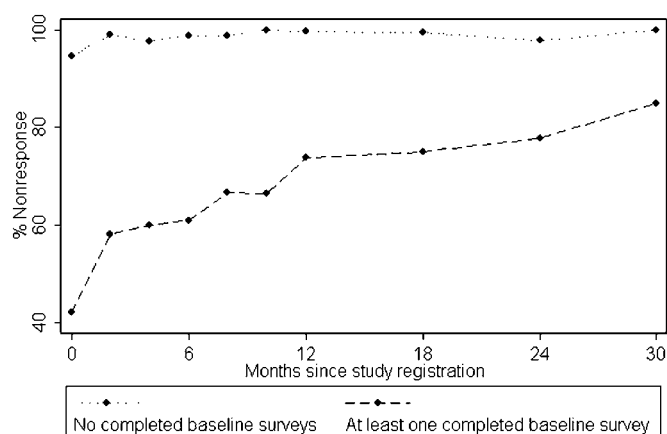


Figure 2 Non-response rate by time since registration.

Another depiction of the effect of time enrolled in IAN on both response rate and time to response can be seen in figure 1. That is, the Weight and Height Survey, which was launched earlier in the protocol, has the highest and quickest response rate compared with the other two surveys that were solicited to those enrolled in the protocol longer (see table 1 for details about time of exposure to survey and time since registration with IAN). Another important illustration in figure 1 is the effect of IMR intervention. This is shown through the vertical increases in the Kaplan–Meier curves for the Weight and Height and Access to Healthcare Surveys, but not the Vaccination History Survey. These increases are a product of email reminders sent to non-responders for the former two surveys, but not the latter. Taken together, these data suggest time since enrollment is a very important predictor of survey response, and further investigation into the effect of novel interventions and informatics tools (eg, REDCap) on response rates is greatly needed.

It is important to note the strengths and limitations of the present study. Specific strengths include the addition of novel predictors, longitudinal design, large sample size, and minimal, if any, response or information bias, because all participants in IAN were included in the study. The major limitation is the generalizability of findings. This is due to the unique design of the project, which has a significant community component that may increase participation in follow-up studies. Another concern is the lack of validity of the child's ASD diagnosis. Because two studies have demonstrated strong correlations between the parent-reported diagnoses in IAN and clinic-based ASD diagnoses, this is of minimal alarm.^{3 4}

CONCLUSION

This study represents an important step toward an empirical understanding of non-response in longitudinal, survey-based IMR. In sum, we found demographic characteristics, time since registration, and participation at baseline all play very important and unique roles in online survey response rates. It is our hope that these results will spawn further investigation into this topic with an eye toward developing novel interventions aimed at assuaging this hazard, which diminishes the quality and interpretability of all research, regardless of design.

Acknowledgments The authors wish to acknowledge the families participating in the Interactive Autism Network that made this study possible. They also thank Connie Anderson, Kiely Law, Alison Marvin, Jay Nestle, and Eleeshabah Yahudah for the work that they do in support of the Interactive Autism Network's research efforts.

Contributors PL is principal investigator for the Interactive Autism Network (IAN). As such, he was responsible for collection of the data employed in the present study. He also provided the original conception of the study. CC contributed to the collection of data and design of instruments within IAN. LK and PL were both responsible for the data analyses. LK created the first draft of the manuscript. HL contributed to the manuscript through assistance with crafting of research questions, critical edits, and review of results from the statistical analyses and figures. All authors contributed to the design of the study and provided substantial edits to the current manuscript.

Funding This study was funded by Autism Speaks Foundation, Simons Foundation, Kennedy Krieger Institute, and the National Institute of Mental Health as part of IAN Core Activity grants. The authors have no financial disclosures.

Competing interests None.

Ethics approval Ethics approval was provided by Johns Hopkins Medicine Institutional Review Board (JHM IRB#: NA_00002750).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Smith A.** Home broadband 2010. *Pew Internet & American Life Project*. <http://pewinternet.org/Reports/2010/Home-Broadband-2010.aspx> (accessed 10 Dec 2011).

2. **U.S. Department of Commerce.** *Exploring The Digital Nation: Computer And Internet Use At Home*. <http://www.esa.doc.gov/sites/default/files/reports/documents/exploringthedigitalnation-computerandinternetuseathome.pdf> (accessed 10 Dec 2011).
3. **Lee H, Marvin AR, Watson T, et al.** Accuracy of phenotyping of autistic children based on Internet implemented parent report. *Am J Med Genet B Neuropsychiatr Genet* 2010;**6**:1119–26.
4. **Daniels AM, Rosenberg RE, Anderson, et al.** Verification of parent-report of child autism spectrum disorder diagnosis to a web-based autism registry. *J Autism Dev Disord* 2012;**42**:257–65.
5. **Smith M, Leigh B.** Virtual subjects: using the Internet as an alternative source of subjects and research environment. *Behav Res Methods Instrum Comput* 1997;**29**:496–505.
6. **Krantz J, Ballard J, Scher J.** Comparing the results of laboratory and world-wide-web samples on the determinants of female attractiveness. *Behav Res Methods Instrum Comput* 1997;**29**:264–9.
7. **Stanton D, Foreman N, Wilson PN.** Uses of virtual reality in clinical training: developing the spatial skills of children with mobility impairments. *Stud Health Technol Inform* 1998;**58**:219–32.
8. **Bordia P.** Studying verbal interaction on the internet: the case of rumor transmission research. *Behav Res Methods Instrum Comput* 1996;**28**:149–51.
9. **Szabo A, Frenkl R.** Consideration of research on the internet: guidelines and implications for human movement studies. *Clin Kinesiol* 1996;**50**:58–65.
10. **Sheehan KB, McMillan SJ.** Response variation in e-mail surveys: an exploration. *J Advertising Res* 1999;**39**:45–54.
11. **Couper MP.** Issues of representation in eHealth research (with a focus on web surveys). *Am J Prev Med* 2007;**32**:S83–9.
12. **Balter KA, Balter O, Fondell E, et al.** Web-based and mailed questionnaires: a comparison of response rates and compliance. *Epidemiology* 2005;**16**:577–9.
13. **Kaplowitz MD, Hadlock TD, Levine R.** A comparison of web and mail survey response rates. *Public Opin Q* 2004;**68**:98–101.
14. **McCabe SE, Boyd CJ, Couper MP, et al.** Mode effects for collecting alcohol and other drug use data: web and U.S. mail. *J Stud Alcohol* 2002;**63**:755–61.
15. **Tse A, Tse K, Yin C, et al.** Comparing two methods of sending out questionnaires: e-mail versus mail. *J Mark Res Soc* 1995;**37**:441–6.
16. **Kiesler S, Sproul L.** Response effects in the electronic survey. *Public Opin Q* 1986;**50**:402–13.
17. **Dodge S, Cucchi P.** The use of an on-line survey to assess visitor satisfaction with a poison control website. *J Toxicol Clin Toxicol* 2002;**40**:629.
18. **Cook C, Heath F, Thompson RL.** A meta-analysis of response rates in web- and internet-based surveys. *Educ Psychol Meas* 2000;**60**:821–36.
19. **Manfreda KL, Bosnjak M, Berzelak J, et al.** Web surveys versus other survey modes: a meta-analysis comparing response rates. *Int J Market Res* 2008;**50**:79–104.
20. **Greenlaw C, Brown-Welty S.** A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost. *Eval Rev* 2009;**33**:464–80.
21. **Glasgow RE, Nelson CC, Kearney KA, et al.** Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. *J Med Internet Res* 2007;**9**:e11.
22. **Jain A, Ross MW.** Predictors of drop-out in an Internet study of men who have sex with men. *Cyberpsychol Behav* 2008;**11**:583–6.
23. **Neil AL, Batterham P, Christensen H, et al.** Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. *J Med Internet Res* 2009;**11**:e6.
24. **McCoy TP, Ip EH, Blocker JN, et al.** Attrition Bias in a U.S. Internet survey of alcohol use among college freshmen. *J Stud Alcohol Drugs* 2009;**70**:606–14.
25. **Stevens VJ, Funk KL, Brantley PJ, et al.** Design and implementation of an interactive website to support long-term maintenance of weight loss. *J Med Internet Res* 2008;**10**:e1.
26. **Joinson A, Paine C, Buchanan T, et al.** Measuring self-disclosure online: blurring and non-response to sensitive items in web-based surveys. *Comput Hum Behav* 2008;**24**:2158–71.
27. **O'Neil KM, Penrod SD.** Methodological variables in Web-based research that may affect results: sample type, monetary incentives, and personal information. *Behav Res Methods Instrum Comput* 2001;**33**:226–33.
28. **Edwards P, Roberts I, Clarke M, et al.** Increasing response rates to postal questionnaires: systematic review. *BMJ* 2002;**324**:118.
29. **Sax L, Gilmartin SK, Bryant A.** Assessing response rates and nonresponse bias in web and paper surveys. *Res Higher Educ* 2003;**44**:409–32.
30. **Khosropour CM, Sullivan PS.** Risk of disclosure of participating in an internet-based HIV behavioural risk study of men who have sex with men. *J Med Ethics* 2011;**37**:768–9.
31. **Rutter M, Bailey A, Lord C.** *The Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services, 2003.
32. **Constantino JN, Gruber CP.** *The Social Responsiveness Scale Manual*. Los Angeles, CA: Western Psychological Services, 2005.
33. **Centers for Disease Control and Prevention.** 2006 NCHS Urban-Rural Classification Scheme For Counties. http://www.cdc.gov/nchs/data_access/urban_rural.htm (accessed 10 Dec 2011).
34. **Rogers WH.** Regression standard errors in clustered samples. *Reprinted Stata Tech Bull Reprints* 1993;**3**:88–94.