Large national datasets assessing health-related use of the internet, technology, and social media

Molly E. Waring, PhD
Assistant Professor, Department of Allied Health Sciences
Director of Methodology Core, UConn Center for mHealth and Social Media
Email molly.waring@uconn.edu
Twitter @DrMollyWaring
6 April 2020

Sampling from a population

• To answer research questions, ideally we would observe all members of the population of interest
• Since this isn’t feasible (practically, ethically, financially), we study a representative sample of the population
  – Sampling unit = entity selected for study
  – e.g., individual, dyad, household, medical visit
• With statistical inference, conclusions based on the sample can be attributed back to the population
Sample weights: the basic idea

• The statistical weight of a sampled person is the number of people in the population that this person represents

• If we surveyed all members of the population
  – Each person represents 1 person
  – Each person has a weight of $1/1 = 1$

• Simple random sample with sampling rate = $1/200$
  – Each sampled person represents 200 people
  – Each sampled person has a weight of $1/200$

More complex sampling

• In more complex sample
  – Sampling weight = probability of selection
    • E.g., $P(\text{select county}) \times P(\text{select segment}) \times P(\text{select household}) \times P(\text{select individual})$
  – Each sampled person represents $1/\text{weight}$ people

• Sampling weights may also be adjusted for non-response or adjusted to represent population characteristics

Weighting analyses to generate estimates representative of the population

• Using analytic techniques that incorporate sample weights allow us to generate results that are representative of the underlying population
• Results from appropriately weighted analyses can be nationally-representative
  – E.g., “x% of US adults”
• Results are estimates, report SE or 95% CI
  – E.g., “50.7% (SE: 2.5%) of US adults…”
  – Different from 50.7% of a sample (exact percent)

Ethical considerations

• IRB review required?
  – For original data collection, absolutely.
  – For research using the de-identified, public-use datasets, no.
  – For research using limited-access datasets that include identifiers or sensitive data, typically yes.
Overview of existing datasets

- Adults: HINTS, Pew, NHIS, CPS
- Children/adolescents: FLASHE, YRBS, NSCH
- Posts/searches: Twitter, Google trends

- [https://hints.cancer.gov](https://hints.cancer.gov)
- HINTS was designed to assess the impact of the health information environment
  - How people access and use health information
  - The degree to which people are engaged in healthy behaviors
- American public's access to and use of info about cancer from prevention, early detection, diagnosis, treatment, and survivorship
  - Including cancer risk perceptions, communication, risk behaviors, understanding of cancer prevention messages
**Timeframe**

- HINTS 1 = 2003 (n=6369)
- HINTS 2 = 2005 (n=5586)
- HINTS 3 = 2008 (n=7674)
- HINTS Puerto Rico = 2009 (n=639)
- HINTS 4
  - Cycle 1 = 2011 (n=3959)
  - Cycle 2 = 2012 (n=3630)
  - Cycle 3 = 2013 (n=3185)
  - Cycle 4 = 2014 (n=3677)
- HINTS FDA
  - Cycle 1 = 2015 (n=3787)
  - Cycle 2 = 2017 (n=1736)
- HINTS 5
  - Cycle 1 = 2017 (n=3285)
  - Cycle 2 = 2018 (n=3504)
  - Cycle 3 = 2019 (n=5438)
  - Cycle 4 = 2020 (planned)


**Design, Sampling, & Generalizability**

- HINTS is a mail survey based on a stratified probability sample of the US adult, civilian, non-institutionalized population
  - Stratification: addresses in areas with high vs low concentrations of minority populations; oversample high-minority areas
  - Randomly selected addresses, adult within households
  - A Spanish version of the questionnaire was distributed to households that had a Hispanic surname
- Weighted analyses are generalizable to Civilian, non-institutionalized US adults
  - Can compare rural vs. urban metropolitan statistical areas (MSAs), census regions (4), and census divisions (9)

Data collection

- Content varies year to year based on trending areas/recent developments in cancer communication
- Ex: HINTS 5 Cycle 2 includes questions on caregiving, palliative care, and family cancer history

Search questions/topics: https://hints.cancer.gov/view-questions-topics/all-hints-questions.aspx

Questions of interest available in HINTS 5 Cycle 2 (2018) and Cycle 3 (2019)

- Internet use; how accessed (e.g., wifi, broadband; home, work)
- Online activities: online support group, watch health-related video on YouTube, social networking site like Facebook or LinkedIn, share health-related info on social media sites, communicate with doctor, seek health information
- Info-seeking: sources, trust in sources
- Have health-related apps
- Use of electronic monitor/device to track health or activity such as Fitbit, BP monitor, blood glucose monitor; willingness to share info with doctor; willingness to share info with family or friends
Example studies

  [link](https://link.springer.com/article/10.1007%2Fs13187-019-01597-0)
- Alcala2019: Trust in Sources of Tobacco Health Information, Perceptions of Harm, and Use of E-Cigarettes
  [link](https://www.ncbi.nlm.nih.gov/pubmed/30715455)
- Bangerter2019: Health Information–Seeking Behaviors of Family Caregivers
  [link](https://aging.jmir.org/2019/1/e11237/)
- Greenberg-Worisek2019: Tracking Healthy People 2020 Internet, Broadband, and Mobile Device Access Goals: An Update Using Data from HINTS
  [link](https://www.jmir.org/2019/6/e13300/)
- Mahmood2019: Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions
  [link](https://www.ncbi.nlm.nih.gov/pubmed/31656632)

Accessing HINTS data – quick stats

In the last 12 months, have you used the Internet for any of the following reasons? Participated in an on-line support group for people with a similar health or medical issue?

<table>
<thead>
<tr>
<th></th>
<th>ESTIMATED US ADULT POPULATION</th>
<th>SURVEY RESPONDENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Response</td>
<td>Number</td>
</tr>
<tr>
<td>1</td>
<td>Yes</td>
<td>13,663,167</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>229,458,262</td>
</tr>
<tr>
<td>9</td>
<td>Missing data (not ascertained)</td>
<td>6,168,343</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>100%</td>
</tr>
</tbody>
</table>

Green = cycles that included the question – click on cycle to bring up responses

Sample n (%) and weighted N (%) – useful for assessing feasibility of study ideas and quick statistics

[link](https://hints.cancer.gov/view-questions-topics/all-hints-questions.aspx)
Downloading HINTS datasets

- [https://hints.cancer.gov/data/download-data.aspx](https://hints.cancer.gov/data/download-data.aspx)
- Must agree to terms of use
- Zip file includes data, formats, and documentation including survey annotated with skip patterns
- SAS, Stata, and SPSS versions available

Strengths & Limitations

- Data on cancer-related risk perceptions, communication, information-seeking, knowledge, and screening/prevention
- No/limited data on other health topics
- Can examine trends across time
- Questions asked vary across cycles
- Cross-sectional
- Data self-reported
- Low response rate (32% in 2017, 33% in 2018)
  - Sample weights do adjust for non-response
• [https://www.pewresearch.org/internet/](https://www.pewresearch.org/internet/)
  • Internet, science, & tech research – focus on how science and technology changes affect families, communities, education, health care and medicine, civic and political life, and workers’ activities
    – Plus lots of other topics! [https://www.pewresearch.org/topics/](https://www.pewresearch.org/topics/)
• RDD sample of landlines and cell phones
• Results representative of US adult population


**Questions of Interest**

• Use of social media platforms, frequency of use
• Broadband at home vs smartphone access
• Ownership of desktop computers, laptops, tablets, e-readers, smartphones, cell phones
• Digital knowledge
• Perceptions of Facebook policies, cybersecurity
• Thoughts about future of technology
• ... and many other questions related to use of the Internet, technology, and social media
Accessing data

- [https://www.pewresearch.org/internet/](https://www.pewresearch.org/internet/)
- Fact Sheets
  - E.g., Mobile, Social Media, Internet/Broadband
- Reports
  - E.g., Americans and Digital Knowledge, Virtues and Downsides of Online Dating
- Most survey datasets available to download
  - Need to create account

Considerations

- Focus on prevalence of activities and differences by demographics (age, gender, race/ethnicity, education, income, rural/suburban/urban)
- Reports examine unadjusted differences between groups
- Typically little other health information collected
National Health Interview Survey (NHIS)

- [https://www.cdc.gov/nchs/nhis/index.htm](https://www.cdc.gov/nchs/nhis/index.htm)
- Nation’s largest in-person household health survey (2018: 25,417 adults and 8,269 children)
- Conducted continuously since 1957
- Collects information on health status and conditions, disability, access to and use of health services, health insurance coverage, immunizations, risk factors, and health-related behaviors
- Generalizable to civilian noninstitutionalized US population

**Core Questionnaires**

<table>
<thead>
<tr>
<th><strong>Table 10</strong>: Conditions and reference periods in the 2018 National Health Interview Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
</tr>
<tr>
<td>Hypertension</td>
</tr>
<tr>
<td>Chronic heart disease</td>
</tr>
<tr>
<td>Asthma</td>
</tr>
<tr>
<td>Cigarette smoking</td>
</tr>
<tr>
<td>Leisure-time PA</td>
</tr>
<tr>
<td>Alcohol use</td>
</tr>
<tr>
<td>Height &amp; weight</td>
</tr>
</tbody>
</table>

### Supplemental Surveys

**Questions of Interest**

<table>
<thead>
<tr>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complementary Health</td>
<td>Asthma</td>
</tr>
<tr>
<td>Cultural Competence (CLAS)</td>
<td>Cancer Screening</td>
</tr>
<tr>
<td>Heart Disease and Stroke</td>
<td>Cognitive Disability</td>
</tr>
<tr>
<td>Cognitive Disability</td>
<td>Food Security</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>Immunization</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Heart Disease and Stroke Prevention (Million Hearts)</td>
</tr>
<tr>
<td>Vision</td>
<td>Tobacco and E-cigarette Use</td>
</tr>
<tr>
<td>Chronic Pain</td>
<td>Internet and Email Usage</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>Disability</td>
</tr>
<tr>
<td>Expanded Content on Health Care Access and Utilization</td>
<td></td>
</tr>
<tr>
<td>Food Security</td>
<td></td>
</tr>
<tr>
<td>Child Mental Health (Brief Strengths and Difficulties Questionnaire)</td>
<td></td>
</tr>
<tr>
<td>Immunization</td>
<td></td>
</tr>
<tr>
<td>Heart Disease and Stroke Prevention (Million Hearts)</td>
<td></td>
</tr>
<tr>
<td>Tobacco and E-cigarette Use</td>
<td></td>
</tr>
<tr>
<td>Internet and Email Usage</td>
<td></td>
</tr>
<tr>
<td>Disability</td>
<td></td>
</tr>
</tbody>
</table>

### 2018 Adult Questionnaire

- Use Internet, email
- Used computer to look online for health information, fill prescription, schedule medical appointment, communicate with health care provider

*** different questions each year

[https://www.cdc.gov/nchs/data/nhis/NHIS_Supplements_and_Sponsors.pdf](https://www.cdc.gov/nchs/data/nhis/NHIS_Supplements_and_Sponsors.pdf)

### Example studies

- Hung2020: Health information technology use among older adults in the US (NHIS 2009-2018)  


- Gonzalez2019: Web-Based Health Information Technology: Access Among Latinos Varies by Subgroup Affiliation (NHIS 2015-2016)  

- Zhang2017: Comparison of Health Information Technology Use Between American Adults With and Without Chronic Health Conditions (NHIS 2012)  
  [https://www.jmir.org/2017/10/e335/](https://www.jmir.org/2017/10/e335/)
**Accessing NHIS data**

- Data briefs  
  [https://www.cdc.gov/nchs/products/databriefs.htm](https://www.cdc.gov/nchs/products/databriefs.htm)
- QuickStats  
  [https://www.cdc.gov/nchs/nhis/nhis_quickstats.htm](https://www.cdc.gov/nchs/nhis/nhis_quickstats.htm)
- Early Release Measures  
  [https://www.cdc.gov/nchs/nhis/releases/released201905.htm](https://www.cdc.gov/nchs/nhis/releases/released201905.htm)
- Data to download  
- 2018 data:  

**Considerations**

- Cross-sectional
- Self-reported data
- Data on health insurance, medical conditions, health care utilization, household composition
- Large sample
- Nationally representative
- Household data, parent-child dyads
- Can combine across years
- Linkages with other datasets
Current Population Survey (CPS)

• Sponsored jointly by the Census Bureau and Bureau of Labor Statistics
  - https://www.census.gov/programs-surveys/cps.html
  - https://www.bls.gov/cps/

• Primary source of labor force statistics for the US
  - Labor force, employment, unemployment, persons not in the work force, hours of work, earnings, demographics

• Probability sample of 60,000 households
  - Designed to produce national and state estimates of labor force characteristics in civilian non-institutionalized population 16+ years

• Supplemental surveys: food security, tobacco use, fertility, computers & Internet use
  - https://www.census.gov/programs-surveys/cps/about/supplemental-surveys.html

Questions of interest

• Use of desktop computer, laptop, tablet, wearable devices; wifi, broadband, dial-up
• Internet use at home, work, school, café, on public transit, library/other community location
• Use email, texting/instant messaging, social media, watch videos, radio/stream music
• Connected household devices (e.g., thermostat)
• Communicate with healthcare provider, access medical records
• Concerns about online privacy, cyberbullying
• Reasons not online at home

https://www2.census.gov/programs-surveys/cps/techdocs/cpsnov17.pdf
Example Publications

- Wright2009: Prescription for trouble: Medicare Part D and patterns of computer and internet access among the elderly
- National Telecommunications and Information Association, 2014: Exploring the Digital Nation: Embracing the Mobile Internet

Accessing CPS data

- Data available via Census and BLS websites
  https://www.bls.gov/cps/data.htm
- Table creator
  https://www.census.gov/cps/data/cpstablecreator.html
  – In progress of migrating to new tool
- DataFerrett https://dataferrett.census.gov/#
  – Data analysis and extraction tool to customize federal, state, and local data to suit your requirements
  – Migration to new tool in progress
  https://data.census.gov/mdat/#/
- Micro data
  https://thedataweb.rm.census.gov/ftp/cps_ftp.html#
Considerations

- Large sample size
- Focus on employment and demographics
- Limited health data

FLASHE
Family Life, Activity, Sun, Health, and Eating Study

- Survey data of psychosocial, generational (parent-adolescent), and environmental correlates of cancer-preventive behaviors
- Cross-sectional Internet-based survey in 2014
- Adolescent and parent/caregiver completed 2 web surveys (n=1479 dyads)
- Balanced sampling: sample to match US households with 1+ 12-17yo as closely as possible on gender, Census division, household income, household size, and race/ethnicity

Data collection

- Teen diet survey
  - Attitudes & opinions
  - What you eat and drink
  - Food away from home
  - Food in your home
  - Family meals
  - Your preferences
  - Your parents

- Teen physical activity survey
  - Physical activity
  - Your home and neighborhood
  - Using electronic devices
  - Time spent in the sun and indoor tanning
  - Tobacco use
  - Sleep
  - Goals in life
  - Your parents

Parent/caregiver surveys – similar topics/questions


Questions of Interest

- Teen & Parent
  - Hours/day use computer, phone, TV, gaming devices, electronic reader in school and outside of school
  - Why would limit use of electronic devices, confidence in ability to limit use

- Parent
  - Teen’s use of computer, phone, TV, gaming devices, electronic reader
  - Parenting rules about screen time

Example studies

• Zhang2019: parenting styles and adolescents’ energy balance-related behaviors

• Joyal-Desmarais2019: Interpersonal effects of parents and adolescents on each other's health behaviours

• Yang2019: The Effect of Screen Viewing Duration and Self-Efficacy in Limiting Screen Viewing on Loneliness in Adolescent-Parent Dyads

• Rice2019: Interactions among perceived norms and attitudes about health-related behaviors in U.S. adolescents

Accessing FLASHE data

• https://cancercontrol.cancer.gov/brp/hbrb/flashe-files.aspx

• SAS code and webinar on how to conduct dyadic analyses
Considerations

• Parent/caregiver-adolescent dyads
• Cross-sectional
• Data self-reported
  – Except for adolescent accelerometer data
• Neat data on cancer risk factors: behaviors, motivations, opinions, parenting rules
• Non-probability sample from volunteer online consumer opinion panel – similar characteristics to US households with adolescents 12-17 years old
  – Parent/caregivers mostly female, highly educated

Youth Risk Behavior Survey (YRBS)

• https://www.cdc.gov/healthyyouth/data/yrbs/index.htm
• National school-based survey conducted by CDC and state, territorial, tribal, and local surveys conducted by state, territorial, and local education and health agencies and tribal governments
• Conducted every 2 years – data available for 1991-2017
• Datasets available: national, by state, large school districts
• Representative of 9th-12th grade students in public and private schools in the US
Data Collection

- Monitors 6 categories of health-related behaviors
- Survey instruments
  https://www.cdc.gov/healthyyouth/data/yrbs/questionnaires.htm
- Questions of interest
  - Texting or emailing while driving
  - Video games/texting/social media use
  - Cyberbullying

Example studies

  https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6119526/
- Alhajji2019: Cyberbullying, Mental Health, and Violence in Adolescents and Associations With Sex and Race (YRBS 2015)
- Baiden2019: The association between excessive screen-time behaviors and insufficient sleep among adolescents (YRBS 2017)
Accessing Data

- Online data analysis tool
  - All US or by state, sex, race/ethnicity, sexual orientation, grade
  [https://www.cdc.gov/healthyyouth/data/yrbs/data.htm](https://www.cdc.gov/healthyyouth/data/yrbs/data.htm)
- Results [https://www.cdc.gov/healthyyouth/data/yrbs/results.htm](https://www.cdc.gov/healthyyouth/data/yrbs/results.htm)
- Infographics [https://www.cdc.gov/healthyyouth/data/yrbs/toolkit.htm](https://www.cdc.gov/healthyyouth/data/yrbs/toolkit.htm)
- Reports of 2017 YRBS

### Played Video or Computer Games Or Used a Computer For 3 Or More Hours Per Day

<table>
<thead>
<tr>
<th>Year</th>
<th>Grade</th>
<th>Total</th>
<th>9th</th>
<th>10th</th>
<th>11th</th>
<th>12th</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>43.0 (41.1-44.9)</td>
<td>45.0 (41.7-48.2)</td>
<td>45.1 (42.3-47.8)</td>
<td>42.3 (38.5-46.1)</td>
<td>39.2 (36.7-41.8)</td>
<td></td>
</tr>
</tbody>
</table>

United States, High School Youth Risk Behavior Survey, 2017

![National Survey of Children's Health](https://www.childhealthdata.org/learn-about-the-nsch/NSCH)

- [https://mchb.hrsa.gov/data/national-surveys](https://mchb.hrsa.gov/data/national-surveys)
- Designed to produce national and state-level data on the physical & emotional health of US children 0-17 years
  - Related factors: medical home, family interactions, parental health, school experiences, safe neighborhoods
- Non-institutionalized children 0-17 years sampled from all 50 states and Washington DC
- Adult who knows the child best (usually a parent) completes online or paper survey about the child
  - 2016: 50,212 nationally, 638-1,351 per state
  - 2017: 21,599 nationally, 343-454 per state

Data collection

- SECTION 1: Initial Screener
- SECTION 2: Survey Questions
  - A. This Child’s Health
  - B. This Child as an Infant
  - C. Health Care Services
  - D. Experience with This Child’s Health Care Providers
  - E. This Child’s Health Insurance Coverage
  - F. Providing for This Child’s Health
  - G. This Child’s Learning (0-5 years)
  - G. This Child’s Schooling and Activities (6-17 years)
  - H. About You and This Child
  - I. About Your Family and Household
  - J. About You
  - K. Household Information

Questions of interest

- On an average weekday, about how much time does this child usually spend:
- … in front of a TV watching TV programs, videos, or playing video games?
- …with computers, cell phones, handheld video games, and other electronic devices, doing things other than schoolwork?

Accessing data

- Interactive data query
  - [http://childhealthdata.org/browse/survey](http://childhealthdata.org/browse/survey)
  - Survey year, US vs state-level
  - Sample N, weighted N, weighted % (95% CI) provided

- Public-use datasets
  - Includes variable list, methodology report, analytic guides
  - Note, major changes to survey administration in 2016; pre-2016 vs 2016+ surveys not comparable
**Example studies**

- **Lo2015**: TV in bedroom and weekday screen time in kids with ADD/ADHD (NSCH 2007)  
- **Sisson2009**: Profiles of sedentary behavior in children and adolescents (NSCH 2001-2006)  

**Twitter**

- **Public tweets**
- Examine what people are saying about a particular health-related topic
- Results generalizable to people who tweet about that topic using keywords/hashtags sampled

Check out our previous webinar: "Methods for Capturing and Examining Social Media Data for Health Research"  
[https://www.youtube.com/watch?v=B-wMT151Nry](https://www.youtube.com/watch?v=B-wMT151Nry)
Accessing Twitter data

• Data collection tools
  – NCapture, NVivo add-on for Chrome
    https://www.qsrinternational.com/nvivo/support-overview/faqs/what-is-ncapture
  – R https://rtweet.info/
  – For-free services (e.g., Gnip)

• Sample posts by hashtag, search term, or user

• Considerations
  – Sampling methods
  – Have to authorize through your social media account
  – Protect user privacy

Example studies

• Kim2020: At the speed of Juul: measuring the Twitter conversation related to ENDS and Juul across space and time (2017-2018)
  https://www.ncbi.nlm.nih.gov/pubmed/32198278

• Griffis2020: Using Social Media to Track Geographic Variability in Language About Diabetes: Analysis of Diabetes-Related Tweets Across the US
  https://diabetes.jmir.org/2020/1/e14431/

• Merrill2020: Posting Post-Blackout: A Qualitative Examination of the Positive and Negative Valence of Tweets Posted After "Blackout" Drinking
• https://trends.google.com/trends
• Google search patterns January 2004 to present
• Search by location worldwide, date, categories, or type of Google search
• RSV = relative search volume
• Accessing data
  – Download data as CSV file
  – Create graphs/charts

Considerations

• Represents Google searches on a topic during time period in geographic area
  – Searches, not people
  – Doesn't include other search engines or other ways to looking for (health-related) information online
• Measure of public interest in topic
• Selecting search terms used to describe concept
• Long-term or seasonal trends
• Compare across geographic region
Example studies

- Niforatos2019: public interest in gun control in the USA
- Caputi2018: Google Searches for “Cheap Cigarettes” Spike at Tax Increases: Evidence from an Algorithm to Detect Spikes in Time Series Data
  https://academic.oup.com/ntr/article/20/6/779/3884451
- Seth2018: Association of socioeconomic and geographic factors with Google trends for tanning and sunscreen
- Madden2017: The seasonal periodicity of health contemplations about exercise and weight loss: ecological correlational study
  https://publichealth.jmir.org/2017/4/e92/
- Jo2015: US consumer interest in non-cigarette tobacco products spikes around the 2009 federal tobacco tax increase

Overview of existing datasets

- Adults: HINTS, Pew, NHIS, CPS
- Children/adolescents: FLASHE, YRBS, NSCH
- Posts/searches: Twitter, Google trends
Interested in learning more about research using large population-based datasets?

- AH6015: Use of Large Population-Based Datasets for Health Promotion
  - Fall 2020 Tuesdays 1:30-4pm
- MS and PhD programs in Health Promotion Sciences
  - [https://healthpromotionsciences.uconn.edu/](https://healthpromotionsciences.uconn.edu/)

UConn Center for mHealth and Social Media Annual Conference

- “Building an Evidence Base for Commercially Available Technology” – VIRTUAL this year – 14-15 May 2020
- More info and registration [https://mhealth.inchip.uconn.edu/events/](https://mhealth.inchip.uconn.edu/events/)
- Agenda, speakers, workshops [https://mhealth.inchip.uconn.edu/uconndigital2020/](https://mhealth.inchip.uconn.edu/uconndigital2020/)
- Abstract submissions due Fri 4/10
  - 1-minute video “poster”