Published in final edited form as:

Ecohealth. 2018 June ; 15(2): 290–301. doi:10.1007/s10393-017-1296-z.

Linking time use data to explore health outcomes: Choosing to vaccinate against Influenza

Kevin Berry^{1,2}, Julia E. Anderson^{1,3}, Jude Bayham^{1,4}, and Eli P. Fenichel¹

¹School of Forestry and Environmental Studies, Yale University, New Haven, CT

²Institute of Social and Economic Research, University of Alaska Anchorage, Anchorage, AK

³School of Public Health, Yale University, New Haven, CT

⁴Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, CO

Abstract

To inform public health and medical decision makers with vaccination interventions, a methodology for merging and analyzing detailed activity data and health outcome is presented. The objective is to investigate relationships between individual's activity choices and their decision to receive an influenza vaccination. Data from the Behavioral Risk Factor Surveillance System (BRFSS) is used to predict vaccination rates in the American Time Use Survey (ATUS) data between 2003 and 2013 by using combined socioeconomic and demographic characteristics. The correlations between the extensive (do or not do) and intensive (how much) decision to perform activities and influenza vaccination are further explored. Significant positive and negative correlations were found between several activities and vaccination. For some activities, the sign of the correlation flips when considering either the intensive or the extensive decision. This flip occurs with highly studied activities, like smoking. Correlations between activities and vaccination can provide an additional metric for targeting those least likely to vaccinate. The methodology outlined in this paper can be replicated to explore correlation among actions and other health outcomes.

Keywords

nontraditional vaccination campaigns; American Time Use Survey; Influenza; Behavioral Risk Factor Surveillance System Survey; public health; vaccination

Introduction

Seasonal influenza is a leading cause of serious illness and death in the United States, infecting approximately 20% of the U.S. population (DHHS, 2012). Based on proven effectiveness, the Advisory Committee on Immunization Practices recommends routine and targeted influenza vaccinations as a frontline defense against infection (CDC, 2016a). Influenza immunization programs have resulted in substantial health improvement and it has

been estimated that during the 2013–14 influenza season vaccination prevented roughly 7.2 million cases of illness, 3.1 million doctor visits, and 90,000 hospitalizations (Reed et al., 2014).

Despite evidence of effectiveness, influenza vaccination coverage (i.e., percentage of targeted population vaccinated) remains sub-optimal (Jacob et al., 2016; Ward, 2014). The United States Centers for Disease Control and Prevention estimates that the vaccination rate for adults age 18 or older is 43.6% (CDC, 2016b). Approximately 70% vaccination coverage nationwide is necessary to block transmission and establish herd immunity (DHHS, 2012; Plans-Rubió & Plans-Rubio, 2012). Increasing vaccination coverage from current levels to 70% is projected to reduce illness by an additional 5.9 million cases, 2.3 million medically attended cases, and 42,000 hospitalizations (DHHS, 2012; Reed et al., 2014).

People's day-to-day actions can inform public health and medical professionals about health outcomes, such as one's decision to vaccinate. This valuable information can be used to tailor health messages, interventions, and programs to meet the needs of particular audiences. Many novel techniques are being adopted to gauge how peoples' attitudes, behaviors, and actions influence health. Researchers are now turning to social media to be used in concert with traditional research methods to capture the opinions of individuals that refuse vaccination (Dredze, Broniatowski, Smith, & Hilyard, 2016). Additionally, new computer science research can infer traditional demographics directly from Big Data, such as Facebook and Twitter, allowing alignment between social media and traditional surveys (Culotta, Kumar, & Cutler, 2015; Dredze et al., 2016; Kosinski, Stillwell, & Graepel, 2013). These Big Data sources can go beyond traditional demographics, identifying fine-grained cultural groups, a precision that is both costly and time consuming with typical surveys, yet critical to understanding vaccine refusal (Dredze et al., 2016).

Knowledge of the types of activities negatively correlated with influenza vaccination can help public health and medical professionals design more effective outreach programs. In recent years public health officials have made use of nontraditional influenza vaccination venues to increase uptake (Postema & Breiman, 2000). Non-traditional settings include pharmacies, grocery stores, work places, and schools (Singleton, Poel, Lu, Nichol, & Iwane, 2005). These programs lower private vaccination costs by making vaccination more convenient, provide information highlighting health benefits, and may ultimately reach people with a lower probability of vaccinating otherwise (Lee, Mehrotra, Burns, & Harris, 2009; Uscher-Pines, Harris, Burns, & Mehrotra, 2012). Targeting the location of these nontraditional sites provides another margin to increase overall coverage. Knowledge of how activity patterns relate to influenza vaccination coverage can help inform where non-traditional settings could be strengthened or introduced.

Targeted flu vaccination messages and nontraditional vaccination settings have successfully increased vaccination coverage and have disproportionately reduced flu cases (Ward, 2014). Flu vaccine education through text message and television advertisement increased influenza immunization among pregnant women, the elderly, and teens (Burns, Ring, & Carroll, 2005; Frew, Saint-Victor, Owens, & Omer, 2014; Stockwell et al., 2012). Outreach at sporting and cultural events in Colorado encouraged people to talk to their physicians about the flu shot,

and increased overall flu vaccines rates (Mostow, 2001). Additionally, intervention by healthcare providers has been shown to increase uptake among patients (Frew et al., 2014).

A majority of people receive their flu vaccine in traditional medical settings; many unvaccinated individuals do not regularly visit or have the time to make a trip to a healthcare provider (Lee et al., 2009). Drug stores and supermarkets, including many national chains, have become popular locations to receive the seasonal flu vaccine (Avery & Lariscy, 2014). Distribution of vaccines through these non-traditional settings has successfully increased influenza immunization (Bearden & Holt, 2005; Weitzel & Goode, 1999).

Peoples' activities may not put them in contact with current messaging campaigns, or may not provide them with regular access to non-traditional vaccination locations. Activities may also relate to several factors known to impact vaccination rates, such as contact with health professionals, social pressure, education and information, available time and perceived health status (Bish, Yardley, Nicoll, & Michie, 2011; Frew et al., 2014; Mullahy, 1998). Interventions to increase flu vaccination coverage have largely been determined by identifying the sociodemographic characteristics of those least likely to receive the vaccination (Mullahy, 1998). Studies often look at immunization uptake with respect to age, health status, race, household structure, income, and insurance coverage (Bish et al., 2011; Merrill & Beard, 2009; Mullahy, 1998).

Time-use survey data has been used in research to explain observed human health characteristics. For instance, Zagheni et al. (2008) use ATUS data to estimate age-specific time of exposure to infectious disease. They fit this time-use data to seroprevalence data for varicella obtained from the National Health and Nutritional Examination Surveys and BRFSS. Their methodology of applying time-use data to helps to identify containment strategies for infectious disease, such as chicken pox. They recommend further research on methods to match location and activity to health-related behaviors to improve understanding of what locations should be targeted for specific interventions (e.g., schools, grocery stores, etc.). In another study by Dukhovnov and Zagheni (2015) time-use data is combined with an Eldercare Roster file to map time spent giving and receiving care by sub-groups of people, revealing notable patterns of care. Childcare took up a sizeable amount of time for caregivers, particularly women in their 30's.

The main objective of this paper is to explore additional methods to connect disparate datasets taken from traditional national surveys. This novel technique of using one dataset to extrapolate information onto a new set of information on individuals' activity patterns will be useful to understand individuals' propensities to be immunized.

Data

The data for this study are compiled from two large national data sources that contain information on activity patterns and health outcomes information, including influenza vaccination. Individual time-use data come from the American Time-Use Survey 2003–2013 (ATUS) (BLS, 2012). The ATUS¹ is subsampled from the U.S. Current Population Survey that contains detailed demographic and socioeconomic information about respondents older

than 15 years of age and their family members (including children under 15 years of age). Survey respondents report a 24-hour diary of activities, locations, and accompanying persons for every minute of the day. The ATUS is a stratified random sample that has been used to analyze behavior and infection dynamics (Bayham, Kuminoff, Gunn, & Fenichel, 2015; Berry, Bayham, Meyer, & Fenichel, 2017; Zagheni et al., 2008) but has not been used to explore vaccination behavior.

We use individual health and vaccination data from the Center for Disease Control Behavioral Risk Factor Surveillance System (BRFSS) (CDC, 2013). The BRFSS² survey has been conducted since 1984 with a cell phone survey introduced in 2011. The survey has covered all 50 US states since 2001. The BRFSS survey is also a stratified sample of the population in each state. Individuals in the survey provide socioeconomic and demographic information as well as information about recent influenza vaccination behaviors.

Statistical Analysis

Our objective is to explore the relationship between vaccination propensities reported in the BRFSS data and activity patterns reported in the ATUS data. While we do not observe activity patterns and vaccination status in a single dataset, we observe a rich set of demographic and socioeconomic characteristics on individuals in each dataset. We use these observable characteristics to extrapolate vaccination propensities from the BRFSS and estimate vaccination likelihood for ATUS respondents.

Extrapolating BRFSS vaccination rates on ATUS Datasets

We estimate a logistic regression model of vaccination uptake using the BRFSS data.

$$\Pr(v \mid \mathbf{x}_i) = \frac{1}{1 + e^{-\eta' \mathbf{x}_i}} \quad \text{for } i = 1, \dots, n \quad (1)$$

The dependent variable v takes the value 1 if the individual obtained a flu vaccination (shot or mist) within the flu season. A vector of observable characteristics, $\mathbf{x}$, includes age in years, sex, race, household income range, highest education level achieved, employment status, student status, marital status, number of household children, state of residence, and year. While the BRFSS records other information that may influence a vaccination decision, our goal is to predict the likelihood that an individual in the ATUS is vaccinated. Consequently, we only include variables reported in both datasets, and only BRFSS observations that include the vaccination variable.³

¹Charts and tables of time spent on various activities by demographic characteristics is available online at the ATUS website, <https://www.bls.gov/tus/home.htm>. The website also includes coding lexicons, data dictionaries, and examples of the questionnaires used. Average time spent on various activities is available on the ATUS Tables subpage.

²BRFSS data is available at <https://www.cdc.gov/brfss/about/index.htm>. This site includes survey data, documentation and examples of the questionnaires used.

³Omitting variables from the regression may bias specific parameter estimates. However, our goal is to predict the likelihood of vaccination and not study the determinants of vaccination. Vaccination determinants such as health care/insurance are correlated with income and employment status, which bias the specific coefficient estimates on income and employment status when omitted but should not affect the out-of-sample prediction for the ATUS respondents of interest.

The model is used to predict the probability that individuals in the ATUS vaccinated against the flu. Simply assigning a vaccination probability based on the model neglects the variation inherent in the BRFSS sample. Consequently, any analysis of activity choice as a function of predicted vaccination probability would underestimate the standard error of the vaccination probability estimates.⁴ To account for the sample variation in our prediction, we construct a Monte Carlo (MC) simulation to estimate confidence intervals around the predicted vaccination probabilities for ATUS respondents. The steps of the MC simulation are as follows:

1. Obtain the coefficient estimates, $\hat{\eta}_i$ and model variance, $\hat{\sigma}^2$, from equation (1) based on the BRFSS data.
2. Use the model estimates to construct a normal sampling distribution with mean equal to the predicted vaccination probability for each combination of socioeconomic and demographic characteristics, $\hat{y}_i$ and variance equal to the forecast or prediction variance. Our assumption that the prediction is normally distributed is supported by non-parametric bootstrap.
3. Construct 100 MC datasets that consist of predicted vaccination probability for all ATUS respondents. Each MC sample represents a draw from the distribution described in step 2. These MC datasets are used to investigate the relationship between activity choices and vaccination likelihoods described in the next section.

Vaccination and Activity Choice

A series of regression models are used to examine the relationship between the probability of vaccination and individual engagement in one of 41 activities (see list in Table S1). A logistic regression is used to examine the relationship between the choice to engage in an activity (extensive decision) and the probability of vaccination. A Quasi-Poisson regression is used to examine the relationship between the number of minutes spent on an activity (intensive decision conditional on engaging in the activity) and the probability of vaccination. The Quasi-Poisson model is a generalization of the standard Poisson regression model designed to accommodate survey weights (Lumley, 2014). Incidentally, the Quasi-Poisson model does not restrict the mean to be equal to the variance as in the standard Poisson regression. The following covariates are included in the extensive and intensive regression models to control for socioeconomic and demographic factors explaining activity choices: age in years, sex, annual income range, highest level of education attained, day of the week, month, year, and state of residence. The odds ratios are calculated for marginal (incremental) increases in vaccination probability for each activity in the logistic regression (extensive decision). In the Quasi-Poisson regression (intensive decision), risk ratios are calculated for a 1% increase in vaccination probability.

We assume that any errors in our estimation of the probability of vaccination are unrelated to errors in our analysis of activity decisions. In other words, there is not some common confounder that remains when estimating the probability of vaccination based on

⁴This is analogous to the problem of underestimating the standard error of an imputed variable.

demographic information that also influences the probability that an individual performs an activity.

The ATUS asks respondents to report the location where an activity was conducted. Grocery stores are a common existing nonconventional vaccination site. The analysis is repeated to examine the relationship between time spent in grocery stores and the extensive and intensive activity decisions. Because the activity and location information all come from the ATUS data, there is no matching uncertainty and Monte Carlo simulations were unnecessary. These models examine the intensive and extensive decisions to perform an activity, conditional on the amount of time spent in a grocery store and a vector of covariates. Odds and Risk ratios are again calculated for the extensive and intensive decisions, respectively. Risk ratios are calculated for both the extensive decision to spend time in a grocery store, and for additional minutes give an individual already goes to the grocery store.

All analysis is conducted in R (R Core Team, 2015). The R package *survey* is used to account for the stratified survey design in the BRFSS and ATUS datasets (Lumley, 2014). The vaccination probability model (1) as well as the extensive and intensive decision models are estimated with R package *svyglm*, which facilitates estimation of generalized linear models where the data come from a stratified random sample.

Results

We present the results of the series of logistic and Quasi-Poisson regression models in a graph to illustrate clustering of activities into four categories (Figure 1). The coefficients on vaccination likelihood variable are plotted as x (extensive, logistic) and y (intensive, Quasi-Poisson) coordinates. Extensive decisions represent whether people engage in the activity at all, while intensive choices are the amount of time spent doing the activity, conditional on engaging in it at all.

The upper left quadrant shows negative correlation (odds ratio less than 1) between the propensity to receive an influenza vaccine and engagement in an activity, but a positive correlation between vaccination and the amount of time spent doing the activity. These activities include ‘tobacco and drug use’ and ‘vehicle maintenance and upkeep’. This result suggests that the decision to use tobacco and drugs is negatively correlated with vaccination, but as the use of drugs and tobacco intensifies, the probability of vaccination increases.

The upper right quadrant displays activities correlated with influenza vaccine uptake on both the intensive and extensive margin. Most activities fall in this realm, indicating that these types of activities are associated with receiving an influenza vaccination. These activities are also primarily social activities, which is consistent with the importance of social pressure influencing decisions to receive vaccination (Bish et al., 2011).

The lower right quadrant of Figure 1 shows activities that are positively correlated with influenza vaccination on the extensive margin, but negatively correlated on the intensive margin. Last, the lower left quadrant contains activities both negatively correlated with flu vaccination on the intensive and extensive margin. These activities are ‘work’, ‘job search’,

and ‘personal care’ (Figure 1). This result is consistent with the literature on work and vaccination that suggests that employed individuals face time budget constraints that may reduce the likelihood of vaccination (Mullahy, 1998).

The next set of results focuses on the relationship between extensive and intensive activity decisions and time spent in a grocery store. Grocery shopping is positively correlated with receiving the flu vaccine. An individual who spends an additional minute grocery shopping is 8% more likely to vaccinate against flu (10% more likely when shopping in general). People engaged in the activities ‘shopping’ and ‘grocery shopping’ are more likely to be exposed to flu vaccination messages and the opportunity to vaccinate at the grocery store.

Despite grocery stores being typical non-conventional vaccination sites, the results suggest that more time spent at the grocery store is negatively correlated with the extensive decision to engage in several activities positively correlated with vaccination. In addition, the decision to spend any time in a grocery store is negatively correlated with the intensive decision to spend time on activities in the following categories. Social pressure is believed to contribute to higher vaccination rates (Bish et al., 2011). Social activities including ‘cardio sports & activities’ and ‘volunteerism’, though negatively correlated with grocery shopping, could lead to social pressure. Additionally, ‘study & research’, and ‘education and class’ (negatively correlated with grocery shopping) are activities related to being a more informed individual in general. The evidence suggests that individuals who are more educated are more likely to vaccinate (Mullahy, 1998).

The extensive and intensive decisions to perform interpersonal activities such as, (e.g. ‘socializing’, ‘child care services’, and others) are positively correlated with time spent in a grocery store and vaccination probability. These activities may be more common for non-working primary care providers in two-person households who would be expected to obtain a flu vaccination to protect children. Substantial amounts of time spent on ‘cardio sports & activities’ may also reflect preferences for health.

Figure 2 depicts the correlation between the 41 activity categories listed in table S1. To focus our discussion, we limit ourselves the activities listed in Table 1. Social activities tend to be positively correlated with each other and with time spent in a grocery store and being vaccinated. The exceptions within the statistically significant activities are ‘education & class’ and ‘work’. While time spent on ‘education & class’ is negatively correlated with time spent in grocery stores, it is positively correlated with several activities that are also positively correlated with vaccination. Conversely, work is negatively correlated with almost every other activity. This suggests that time spent working simply reduces the available time budget to become vaccinated, and individuals who work more lack opportunities to perform other activities.

Discussion

The American Time Use Survey and Behavioral Risk Factor Surveillance System Survey can be used to identify innovative approaches to combine existing public health datasets and provide insight into locations where nontraditional vaccination campaigns may be targeted.

This new method of extrapolating information from one disparate data set onto another to target policy involves examining the activity choice of individuals least likely to receive the flu vaccine, rather than socio-demographic characteristics, as previously done.

Previous work has targeted vaccination interventions based on socio-demographic characteristics and physical location (Mullahy, 1998; Scheminske et al., 2015; Stehr-Green, Sprauer, Williams, & Sullivan, 1990). The novel approach of combining datasets allows for the use of detailed activity data and vaccination data to suggest locations for nontraditional interventions using activity choice where unvaccinated individuals are likely to be. By using the ATUS and BRFSS datasets, which have detailed demographic information, it is possible to examine the relationship between expected vaccination probability and activity choice. The methods outlined in this study can also be used to understand how activity choice effects other health outcomes captured by BRFSS.

The correlation analysis of activity choice and the decision to be vaccinated for influenza has potential to provide valuable insights for policymakers. Previous results from the literature, such as the importance of contact with medical professionals, are supported by our findings; where time spent on activities is correlated with a greater likelihood of vaccination (household child medical care has an odds ratio (OR) of 1.28 for the extensive decision, and a risk ratio (RR) of 1.08 for the intensive decision). Time spent on ‘study & research’ (OR: 1.74, RR: 1.11) or ‘education & class’ (OR: 1.68, RR: 1.07) leads to greater vaccination uptake. Similar results are available for ‘socializing’, where higher social pressure may cause greater vaccination (OR: 1.12, RR: 1.08).

The activity ‘smoking’ demonstrates the importance of considering both the extensive and intensive activity decisions. The extensive decision to use tobacco is negatively correlated with vaccination, consistent with the literature (Nichol, Mac Donald, & Hauge, 1996). Smoking also causes individuals to be at greater risk of influenza (Arcavi & Benowitz, 2004). We cross validate our results using logistic regression on only BRFSS data on vaccination choice against extensive choice data on if an individual is a smoker, controlling for age, sex, race, income, education, employment status and year. This analysis yields a small and negative effect (coefficient of -0.429 with SE of $.00289$) with an odds ratio of $.995$ for a 1% increase in the log-odds of smoking. This is compared to an odds ratio of 0.9 for extensive margin effect associated with engaging in ‘tobacco or drug-use’ from the BRFSS-ATUS combined data.

The intensive decision to smoke (how much to smoke conditional on smoking) is positively correlated with the probability of vaccination (1.18 with p-value of 0.1398). The evidence suggests that a healthcare provider’s recommendation to receive a vaccine can overcome the effect of an individual’s own attitudes concerning the vaccine, leading the individual to vaccinate (Nichol et al., 1996). Increased smoking is correlated with negative health impacts, potentially leading to greater exposure to healthcare professionals. The increased risk to heavy smokers of influenza means the vaccine is particularly important for these individuals, suggesting campaigns targeting smokers might also discriminate based on how much an individual smokes (Arcavi & Benowitz, 2004).

Hiking is a second example of the importance of considering both extensive and intensive decisions. On the intensive margin, hiking is negatively correlated with the probability of vaccination (RR: .948). This implies that individuals who are more likely to be vaccinated spend less time hiking, all-else-equal. Time spent in grocery stores is negatively correlated with the choice to spend time hiking on the intensive and extensive margins (OR: .993, RR: .988). Individuals who hike are less likely to spend time in a grocery store and less likely to be vaccinated. Perhaps because they miss nontraditional campaigns, they may perceive themselves as healthier, or maybe less likely to face social pressure to vaccinate.

Limitations

Our analysis does come with several limitations: we do not identify the marginal individual, and we do not establish a causal relationship between vaccination and activity decisions. This exercise is intended to demonstrate how information from two disparate datasets containing some overlapping information can be linked using socioeconomic and demographic characteristics. The more rigorous identification strategy needed to provide evidence of causation is beyond the scope of this paper.

The ATUS has relatively few observations for some geographic locations and activities that are not common (BLS, 2012). This can be a result of sparse population. To avoid individual identity, demographic data may be omitted making extrapolation more difficult. Omitted data can reduce statistical power, and effect the probability of correctly identifying significant correlations. This limitation can be overcome by performing surveys of potentially important activity in locations of interest. For example, policymakers may sponsor a survey in their area of individuals' time use decisions. The results of this survey could be combined using demographic data to other available vaccination or health data to identify correlations between activities, behaviors, and outcomes of interest. In addition, the application of our method is also limited by the availability of other activity choice and behavior data, and the design of these surveys is an entirely separate research topic.

Conclusion

We apply a novel method to extrapolate information from one disparate dataset to another from traditional national surveys. We use these methods to examine activity choices for insights into increasing vaccination rates based on better targeting interventions. For example, smokers who do not smoke enough to suffer from severe enough health consequences to be in regular contact with health professionals should also be targeted through interventions placed outside of traditional areas. Our approach of using time-use data, which is regularly collected by the US government, could be extended to other public health issues in order to provide more information for targeting policy. This same methodology could be applied to vaccination decisions related to pneumonia, HPV, Shingles, MMR, or screenings for cancers and HIV. Activity data may be of interest where activity is directly related to susceptibility to a disease, such as obesity and its associated health impacts. The BRFSS codebooks contain a variety of health-related variables, and an understanding of activity choice may provide insight into where public health professionals should focus their attention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This publication was made possible by grant number 1R01GM100471-01 from the National Institute of General Medical Sciences (NIGMS) at the National Institutes of Health and NSF. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS. This work was also funded by NSF grant no. 1414374 as part of the joint NSF-NIH-USA Ecology and Evolution of Infectious Diseases program

References

- Arcavi L, Benowitz NL. Cigarette smoking and infection. *Archives of Internal Medicine*. 2004; 164(20):2206–2216. [PubMed: 15534156]
- Avery EJ, Lariscy RW. Preventable disease practices among a lower SES, multicultural, nonurban, US Community: the roles of vaccination efficacy and personal constraints. *Health Communication*. 2014; 29(8):826–836. [PubMed: 24266742]
- Bayham J, Kuminoff NV, Gunn Q, Fenichel EP. *Proc R Soc B*. Vol. 282. The Royal Society; 2015. Measured voluntary avoidance behaviour during the 2009 A/H1N1 epidemic; 20150814
- Bearden DT, Holt T. Statewide impact of pharmacist-delivered adult influenza vaccinations. *American Journal of Preventive Medicine*. 2005; 29(5):450–452. [PubMed: 16376709]
- Berry K, Bayham J, Meyer SR, Fenichel EP. The Allocation of Time and Risk of Lyme: A Case of Ecosystem Service Income and Substitution Effects. *Environmental and Resource Economics*. 2017:1–20.
- Bish A, Yardley L, Nicoll A, Michie S. Factors associated with uptake of vaccination against pandemic influenza: a systematic review. *Vaccine*. 2011; 29(38):6472–6484. [PubMed: 21756960]
- BLS. *American Time Use Survey*. Washington, DC: 2012.
- Burns VE, Ring C, Carroll D. Factors influencing influenza vaccination uptake in an elderly, community-based sample. *Vaccine*. 2005; 23(27):3604–3608. [PubMed: 15855020]
- CDC. Behavioral Risk Factor Surveillance System Survey Data. Atlanta, Georgia: 2013. Retrieved from http://www.cdc.gov/brfss/technical_infodata/surveydata.htm
- CDC. ACIP Recommendations. 2016a. Retrieved from <http://www.cdc.gov/vaccines/acip/recs/index.html>
- CDC. Flu Vaccination Coverage, United States, 2014–15 Influenza Season. 2016b. Retrieved from <https://www.cdc.gov/flu/fluview/coverage-1415estimates.htm>
- Culotta A, Kumar NR, Cutler J. Predicting the Demographics of Twitter Users from Website Traffic Data. *Proceedings of the 29th AAAI Conference on Artificial Intelligence*; 2015. 72–78.
- DHHS. *Healthy People 2020*. 2012. Retrieved from www.healthypeople.gov
- Dredze M, Broniatowski DA, Smith MC, Hilyard KM. Understanding Vaccine Refusal: Why We Need Social Media Now. *American Journal of Preventive Medicine*. 2016; 50doi: 10.1016/j.amepre.2015.10.002
- Frew PM, Saint-Victor DS, Owens LE, Omer SB. Socioecological and message framing factors influencing maternal influenza immunization among minority women. *Vaccine*. 2014; 32(15): 1736–1744. [PubMed: 24486366]
- Jacob V, Chattopadhyay SK, Hopkins DP, Morgan JM, Pitan AA, Clymer JM, Force CPST. Increasing Coverage of Appropriate Vaccinations: A Community Guide Systematic Economic Review. *American Journal of Preventive Medicine*. 2016; 50(6):797–808. [PubMed: 26847663]
- Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*. 2013; 110(15):5802–5805. DOI: 10.1073/pnas.1218772110
- Lee BY, Mehrotra A, Burns RM, Harris KM. Alternative vaccination locations: who uses them and can they increase flu vaccination rates? *Vaccine*. 2009; 27(32):4252–4256. [PubMed: 19406181]

- Merrill RM, Beard JD. Influenza vaccination in the United States, 2005–2007. *Medical Science Monitor Basic Research*. 2009; 15(7):PH92–PH100.
- Mostow SR. *International Congress Series*. Vol. 1219. Elsevier; 2001. Use of alternative sites to administer influenza vaccine improves acceptance by both physicians and patients; 703–706.
- Mullahy J. It'll only hurt a second? Microeconomic determinants of who gets flu shots. *National Bureau of Economic Research*; 1998.
- Nichol KL, Mac Donald R, Hauge M. Factors associated with influenza and pneumococcal vaccination behavior among high-risk adults. *Journal of General Internal Medicine*. 1996; 11(11):673–677. [PubMed: 9120653]
- Plans-Rubió P, Plans-Rubio P. The vaccination coverage required to establish herd immunity against influenza viruses. *Preventive Medicine*. 2012; 55(1):72–77. [PubMed: 22414740]
- Postema AS, Breiman RF. Adult Immunization Programs in Nontraditional Settings: Quality Standards and Guidance for Program Evaluation: A Report of the National Vaccine Advisory Committee. *Morbidity and Mortality Weekly Report: Recommendations and Reports*. 2000:vii–13.
- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. Retrieved from <https://www.r-project.org/>
- Reed C, Kim IK, Singleton JA, Chaves SS, Flannery B, Finelli L, ... Jernigan D. Estimated influenza illnesses and hospitalizations averted by vaccination--United States, 2013–14 influenza season. *MMWR. Morbidity and Mortality Weekly Report*. 2014; 63(49):1151–1154. [PubMed: 25503917]
- Scheminske M, Henninger M, Irving SA, Thompson M, Williams J, Shifflett P, ... Naleway AL. The Association Between Influenza Vaccination and Other Preventative Health Behaviors in a Cohort of Pregnant Women. *Health Education & Behavior*. 2015; 42(3):402–408. [PubMed: 25504630]
- Singleton JA, Poel AJ, Lu PJ, Nichol KL, Iwane MK. Where adults reported receiving influenza vaccination in the United States. *American Journal of Infection Control*. 2005; 33(10):563–570. [PubMed: 16330304]
- Stehr-Green PA, Sprauer MA, Williams WW, Sullivan KM. Predictors of vaccination behavior among persons ages 65 years and older. *American Journal of Public Health*. 1990; 80(9):1127–1129. [PubMed: 2382756]
- Stockwell MS, Kharbanda EO, Martinez RA, Vargas CY, Vawdrey DK, Camargo S. Effect of a text messaging intervention on influenza vaccination in an urban, low-income pediatric and adolescent population: a randomized controlled trial. *JAMA*. 2012; 307(16):1702–1708. [PubMed: 22535855]
- Uscher-Pines L, Harris KM, Burns RM, Mehrotra A. The growth of retail clinics in vaccination delivery in the US. *American Journal of Preventive Medicine*. 2012; 43(1):63–66. [PubMed: 22704748]
- Ward CJ. Influenza vaccination campaigns: is an ounce of prevention worth a pound of cure? *American Economic Journal: Applied Economics*. 2014; 6(1):38–72.
- Weitzel KW, Goode JV. Implementation of a pharmacy-based immunization program in a supermarket chain. *Journal of the American Pharmaceutical Association* (1996). 1999; 40(2):252–256.
- Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong J, Edmunds WJ. Using time-use data to parameterize models for the spread of close-contact infectious diseases. *American Journal of Epidemiology*. 2008; 168(9):1082–1090. [PubMed: 18801889]

Appendix

Vaccination and Activity Choice

Two regression models are used to examine the relationship between vaccination and activities at the extensive (logistic regression) and intensive (Poisson regression) margin.

Estimating the do- don't do decision (extensive choices)

The logistic regression of whether or not individuals perform an activity consists of a binary indicator equal to one if respondent $i = 1, \dots, n$ does activity $j = 1, \dots, J$ as a function of the probability of vaccination, $\hat{v}_i$, and the vector of controls, $\mathbf{x}_i$.

$$\Pr(Bin_{-}a_{ij} = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \hat{v}_i + \beta' \mathbf{x}_i)}} \quad (2)$$

When $\Pr(Bin_{-}a_{ij} = 1)$ the individual chooses to participate in an activity, otherwise it equals zero. The estimated vaccination probabilities, $\hat{v}_i$, have associated prediction error. Therefore, Monte Carlo simulation is used to estimate the vaccination coefficient and associated standard errors. The Monte Carlo procedure is as follows:

1. Draw a vaccination probability from a normal distribution with mean equal to the coefficient estimated from the preliminary logistic regression in equation (1), η_v , and its standard error.
2. Estimate the logistic activity regression in equation (2) and store coefficient β_1 .
3. Repeat steps 1 and 2 100 times
4. Calculate the mean and standard error based on the 100⁵ different coefficient estimates from each run.

Odds ratios are calculated for the impact of a 1% increase in vaccination probability on the likelihood of performing an activity. All odds ratios and risk ratios are calculated for a 1% increase in relative risk of vaccination. Odds ratios from equation (2) are interpreted as the odds that an individual participates in activity j , conditional on $\hat{v}_i$ being 1% higher, relative to the odds that an individual participates in j conditional on their vaccination probability being unchanged.

Estimating time spent (duration decision)

The intensive activity decision, conditional on doing the activity, is modelled as a Poisson regression. The minutes spent doing an activity, a_{ij} , is regressed on the continuous probability of being vaccinated $\hat{v}_i$ and a vector of controls $\mathbf{x}_i$.

⁵We chose 100 estimates due to computing time constraints, currently the intensive decision model takes 4.37 hours to estimate, and the extensive decision takes 22.75 hours to estimate.

$$Pr(a = a_{ij} | x_i) = \frac{e^{-\mu_{ij}} \mu_{ij}^{a_{ij}}}{a_{ij}!}, \quad a_{ij} = 0, 1, 2, \dots \quad (3)$$

$$\ln(\mu_{ij}) = \delta_0 + \delta_1 \hat{v}_i + \delta' x_i + \varepsilon_i$$

Which is estimated with quasi-Poisson errors, which assume that variance is a linear function of the mean to control for possible over dispersion. The coefficient, δ_1 , and its standard error are estimated by the Monte Carlo simulation methods described in the previous section.

Risk ratios are calculated for the second equation in (3) and are interpreted as the ratio of the probability an individual spends another minute participating in an activity conditional on their vaccination probability being 1% higher, relative to the probability of spending another moment participating in that activity conditional on their vaccination probability being unchanged.

Grocery Shopping

We also estimate a model of activity choice and grocery store attendance. Because it is not necessary to statistically match the datasets, we use the full ATUS dataset and do not perform Monte Carlo simulations. We observe the duration of each activity done by respondents on their interview day as well as time spent at the grocery store location.

Estimating time spent (duration decision) consists of a logit model of the decision of how much time to spend at the grocery store on the extensive decision to participate in an activity. Let g_i be the total minutes respondent i spends at a grocery store and the binary variable, d_i^g , be equal to one if respondent i spends at least one minute at a grocery store.

$$Pr \left(Bin_{a_{ij}} = 1 \right) = \frac{1}{1 + e^{-(\gamma_0 + \gamma_1 d_i^g + \gamma_2 g_i * d_i^g + \alpha' x_i)}} \quad \forall i, j \quad (3)$$

Variable a_{ij} remains the time spent on activity j by person i and d_i^g is time spent at a grocery store. We include the vector x_i of controls that consist of age, sex, income, education, the day of the week, month and year as well as the state. We again estimate this equation for each activity j in the ATUS. The model contains both the binary grocery store variable and the interaction between the binary and continuous variable. Including both variables provides flexibility to the model to accommodate the difference between the effect of zero and one minute versus 10 and 11 minutes on activity.

We calculate odds ratios for all activities in equation (3) for both the first minute spend in a grocery store, and every additional minute. The interpretation of the odds ratio for γ_1 is the

ratio of the odds that an individual participates in activity j subject to spending any time at all in a grocery store (binary variable, d_i^g) relative to the odds that an individual participates in activity j subject to not spending time in a grocery store. The odds ratio for γ_2 , is the ratio of the odds an individual participates in an activity conditional on spending an additional minute at the grocery store, versus not having spent more time at the grocery store (but still having gone to a grocery store for at least one minute).

We use a count model to study the association between intensive activity decisions and the extensive decision to spend time in grocery store, d_i^g , and the amount of time in grocery store, g_i ,

$$\Pr(a = a_{ij} | x_i) = \frac{e^{-\lambda_{ij}} \lambda_{ij}^{\hat{a}_{ij}}}{\hat{a}_{ij}!}, \quad (4)$$

$$\ln(\lambda_{ij}) = \gamma_0 + \gamma_1 d_i^g + \gamma_2 g_i * d_i^g + \alpha' x_i + \varepsilon_i$$

The model uses quasi-Poisson methods as an *ad hoc* control for over dispersion. The risk ratio for γ_1 and γ_2 are analogous to the odds ratios for γ_1 and γ_2 in the first stage, except they are not ratios of probabilities instead of odds.

Figure 1.

The probability of getting the flu vaccine based on activity is presented below. The x-axis is the probability of getting the flu vaccine subject to engaging in the activity (extensive margin). The y-axis is the probability of getting the flu vaccine subject to an additional minute spent engaging in the activity (intensive margin). From this plot, we can discern activities correlated with the probability of getting the flu vaccine.

Figure 2.

The cell colors represent the value of the pairwise Pearson correlation coefficient estimate from -1 (red) to 1 (blue). A black “x” indicates an insignificant ($\alpha=0.10$) correlation estimate. In cases with insufficient observations to calculate the pairwise correlation, the estimate was set to 0 and considered insignificant.

Results from our regression analysis which are significant at the 5% level on the extensive and intensive margins for both vaccination and grocery store correlations are shown below. Additional results are available in the Appendix

Table 1

	Grocery Store						Vaccination					
	Extensive Activity			Intensive Activity			Extensive Activity		Intensive Activity			
	γ_1 Odds Ratio	p-value	γ_2 Odds Ratio	p-value	γ_1 Odds Ratio	p-value	Odds Ratio	p-value	Risk Ratio	p-value		
Activities												
medical care	1.58	1.75×10^{-9}	0.998	0.0181	0.614	3.59×10^{-15}	1.003	1.88×10^{-8}	1.19	0	1.11	0
socializing	1.33	2.96×10^{-24}	0.999	1.59×10^{-7}	0.779	1.84×10^{-18}	1.001	2.81×10^{-18}	1.12	0	1.08	0
child medical care	1.11	3.08×10^{-3}	1.003	1.35×10^{-24}	0.92	2.67×10^{-2}	1.001	2.86×10^{-7}	1.28	0	1.08	0
child care services	1.25	4.43×10^{-11}	1.002	6.45×10^{-13}	0.835	6.19×10^{-11}	1.001	5.26×10^{-8}	1.21	0	1.13	0
cardio sports & activities	1.47	1.21×10^{-23}	0.998	1.09×10^{-10}	0.905	6.89×10^{-5}	1.002	9.13×10^{-13}	1.07	0	1.07	0
household care	2.59	6.50×10^{-166}	1.002	9.18×10^{-8}	0.822	1.21×10^{-35}	1.001	7.77×10^{-39}	1.14	0	1.10	0
household care services	4.01	1.16×10^{-23}	0.995	2.16×10^{-3}	0.4995	1.36×10^{-7}	1.007	5.88×10^{-7}	1.07	0	1.13	0
study & research	1.02	0.868	0.997	4.31×10^{-3}	0.87	8.83×10^{-2}	1.001	2.75×10^{-2}	1.74	0	1.11	0
education & class	0.81	7.84×10^{-2}	0.997	6.87×10^{-3}	0.751	6.21×10^{-3}	1.002	7.56×10^{-3}	1.68	0	1.075	0
volunteerism	1.46	1.12×10^{-10}	0.997	1.98×10^{-6}	0.75	4.39×10^{-5}	1.002	7.08×10^{-6}	1.13	0	1.07	0
hobbies & leisure	1.54	6.57×10^{-20}	1.001	0.0362	0.95	1.94×10^{-4}	1.001	1.31×10^{-7}	1.11	0	1.11	0
work	0.85	8.81×10^{-6}	0.996	1.70×10^{-28}	0.86	7.98×10^{-25}	0.9995	4.16×10^{-2}	0.38	0	0.98	0