What is the relationship between online activity and driving-licence-holding amongst young adults?

Postprint of:

Le Vine, S., Latinopoulos, C., Polak, J. (2015) What is the relationship between online activity and driving-licence-holding amongst young adults? *Transportation*. http://dx.doi.org/10.1007/s11116-014-9528-3

Scott Le Vine[1]
Research Associate, Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London, Exhibition Road, London, SW7 2AZ, UK
Telephone: +44 20 7594 6105
Fax: +44 20 7594 6102
Email : slevine@imperial.ac.uk

Charilaos Latinopoulos
Doctoral candidate, Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
Telephone: +44 79 2356 6287
Email : charilaos.latinopoulos10@imperial.ac.uk

John Polak
Chairman, Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
Telephone: +44 20 7594 6100
Email: j.polak@imperial.ac.uk

[1] Corresponding author

Abstract

There is growing interest amongst both practitioners and researchers in the correlates of young adults' driving-licence-acquisition. One aspect of the ongoing scholarly debate is whether taking part in online (i.e. virtual) activities may be associated with young adults feeling less need to drive and hence to acquire a driving licence.

This paper addresses this issue by drawing on analysis of two distinct datasets. Both contain rich pseudo-diary instruments in which people indicate detailed characteristics of their unique online-activity profile. This includes both indicators of the *types* of online activities in which respondents participate, and a separate metric of internet-use *intensity*. The latter is defined in one dataset as the amount of time per week spent online, and in the other dataset the frequency of their internet use.

On the basis of a set of multivariate regression analyses, a positive (i.e. complementary) cross-sectional relationship between young adults' online activity and licence-holding was found. We find that young adults who use the internet are, net of confounding effects, more likely to hold a driving licence than otherwise identical young adults who do not use the internet. Both datasets show this type of effect, and it is robust across a range of model specifications, including multi-stage estimations to address cross-correlation between indicators of internet usage. In addition to the positive net statistical association, we also report several other noteworthy effects. Of the six effects associated with online-activity types that are directly comparable between the two datasets, we find that the correlation in the parameter estimates across the two datasets is 0.63. This suggests similar types of relationships across the two datasets. Also, in several (but not all) of our analyses we found an inverted 'U' shaped *ceteris paribus* relationship between intensity-of-internet-use and licence-holding.

The positive net statistical association between internet use and licence-holding is a different relationship than previously reported in the literature, and therefore further research is needed to reconcile the differences (which are likely due at least in part to different methodological approaches and data resources). Further research is also needed to continue to resolve between the relative saliency of other hypothised determinants of licence-holding (e.g. economic and socio-demographic explanators, as well as licence-acquisition regimes that vary by time and place).

Key words

Youth licencing, online activity, internet

1. Introduction

The interrelationship between virtual (online) activities and real-world activities is an area of growing interest in recent years. The net effect of internet use on mobility patterns is however ambiguous (Mokhtarian 1990); both negative (substitution) and positive linkages (complementarity) are plausible. Despite the ambiguity regarding the relative magnitude of these countervailing effects, it is frequently suggested that increased online-connectivity has led people to travel less, particularly by car, and perhaps even to feel less need to acquire a driving licence (Alcindor 2012, Buccholz and Buccholz 2012, Kay et al. 2014).

The present study investigates the possibility of links between internet use and car-driving-licence-holding. Whilst the analysis we report here is cross-sectional (due to data limitations), it is worth noting that in recent years licence-holding was observed to fall for young people in many developed countries (ifmo 2013, Le Vine and Jones 2012, Delbosc and Currie 2013a, Delbosc and Currie 2013b, Sivak and Schoettle 2012, Grimsrud and El-Geneidy 2013), in general more strongly for young men than women.

Some of the standard explanations for this decline in licence acquisition are increasing cost and difficulty of learning to drive along with the deterrence from the costs car ownership and use. Other more speculative hypotheses suggest that there may be a link between the continuing development of telecommunications technologies (particularly ones based on mobile, rather than wired-in, connectivity) and reduced licence-acquisition rates (Sivak and Schoettle 2012).

In this paper we report findings based on analysis of two different datasets: the Scottish Household Survey (SHS) and the Opinions and Lifestyle Survey (OPN) by the UK Office of National Statistics.

The SHS is a nationally-representative dataset with uniquely-attractive properties for the study of mobility and online activity linkages. The SHS uniquely combines mobility-relevant data (from both a household interview and a one-day travel diary) along with a pseudo-diary of online activity participation. The OPN dataset also contains a pseudo-diary of online activity, as well as whether each respondent holds a driving licence.

Our findings are contrary to the most directly comparable previous study of this issue in the scientific literature, which suggested a cross-sectional net *substitution* relationship (Sivak and Schoettle 2012). This discrepancy with earlier findings led us to confirm that our results are replicated when analysing each of the two completely independent datasets (the SHS and OPN datasets), which employed distinct data generation methods from each other and hence have different data properties. Our analyses of both of these cross-sectional datasets suggest that there is a net *complementary* relationship between young adults' licence-holding and their online activity.

The remainder of this paper is structured as follows: Section 2 discusses the background of the virtual/physical activities interrelationship and the hypothesised mechanisms. The empirical analyses from the SHS and OPN datasets are presented in Sections 3 and 4 respectively. Finally, the implications of the findings from this study are discussed in Section 5.

2. Background

Advances in telecommunications technology, and the rapid rate at which costs have decreased, today provide people with the capability to perform a wide array of activities that previously required physical travel to out-of-home locations. Among other benefits, performing activities online may lead to reduced costs, time and inconvenience. Perhaps more importantly, the online world enables one to take part in entirely new types of activities that do not have direct physical world analogues.

2.1 Substitution v. complementarity

The development of cyberspace has coincided with declining levels of car-based mobility and licence-holding among young people, a trend that has been observed across much of the developed world (Kuhnimhof et al. 2012). Young adults also have a higher tendency to participate in online activities and spend more time online compared to people from older age groups (Latinopoulos et al. 2013). It is widely suggested in the literature that these two phenomena are linked and that the rise of the internet as a medium for communications and interaction might be contributing to the decline in young people's mobility (e.g. Sivak and Schoettle 2012, Litman 2012, Puentes 2012, Kay et al. 2014).

By way of contrast, Cao et al (2010) report evidence contrary to this hypothesis, suggesting that online searching and online buying have a complementary effect on in-store shopping. In the same vein, Douma (2003) addresses the issue of shoppers using the internet to browse for products before leaving from home, in order to make their shopping trip more time-efficient, therefore suggesting also a complementary relationship between e-shopping and physical shopping. Mokhtarian (2002) more broadly suggests that while short-run, single-application studies often find substitution between telecommunications and travel, complementarity is likely to dominate in the long-run.

2.2 Other types of linkages

Mokhtarian (2002) suggests that complementarity can be subdivided into two categories: *enhancement,* where one mode of communication directly causes or facilitates the use of the other (e.g. generate a working trip by sending an important e-mail), and *efficiency,* when one of the two modes increases the efficiency of the other (e.g. air traffic control, logistics operations).

Another form of linkage included in this taxonomy is *modification.* In this case the use of one mode modifies something in the process of the other mode's use without either stimulating or dampening it. An example of modification is when a person uses in-vehicle navigation in real-time and alters the initial route of his trip based on the received information.

Finally there is the possibility of *neutrality,* if the use of one mode of communication leaves the other mode unaffected (e.g. taking an online course does not necessarily mean that the individual would otherwise be registered at an institution to physically attend classes).

2.3 Mechanisms possibly linking licence-acquisition and online-activity

As the rate at which young adults acquire driving-licences has declined in many countries, there is an increasing interest in understanding the correlates of young adults' licence-acquisition.

A driving licence can be conceptualised as a type of *mobility tool* (Beige and Axhausen 2006), similar in some ways to owning or otherwise having access to a car itself. Both are prerequisites for car-based mobility, and an emerging body of literature is showing that people's choices of whether to acquire such mobility tools (alternatively referred to as *mobility resources* [Le Vine et al. 2013] or *mobility attributes* (Vovsha and Petersen [2009]) can be meaningfully conceived of as a function of people's perceived needs to use the methods of travel that they enable.

This line of research can be considered in light of the body of theory regarding telecommunications / transportation interactions. It may plausibly be hypothesised that if telecommunications serves as either a net substitute or complement for independent car-based transportation, this could feed through to impacts on people's perceived needs to obtain a driving licence.

Therefore the motivation of the study was to establish whether there is any evidence to support the suggestion of a negative linkage between online-activity and licence-acquisition. The literature on this specific issue can be organised into three classes of studies.

The first type of study employs data from cross-sectional surveys in which young adults without a driving licence are asked to choose the reason(s) that they do not have one from a pre-specified list. Williams (2011) reports that 7% of U.S. teenagers (age 15 – 18) who had not yet started the licence-acquisition process cited *"Facebook, texting, etc. keep me in touch with friends"* as a contributory factor. Based on a nationally-representative survey, Tefft et al. (2013, 2014) report that 17% of U.S. teenagers aged 18 – 20 indicated that the fact that they *"Could connect with friends online"* was at least in part responsible for them deferring licence acquisition. Schoettle and Sivak (2013a) report that 8% of unlicenced U.S. adults aged 18 – 39 indicate that *"Able to communicate and/or conduct business online instead"* was responsible partly or fully for their unlicenced status. These studies employ cross-sectional datasets and purely descriptive statistical analyses.

The second class of study in the literature draws on exploratory qualitative methods. Delbosc and Currie (2014), to the authors' knowledge the only recent example of this type of study in the public literature, report on online discussion groups of young adults living in Melbourne, Australia (n=33 young adults, in 3 discussion groups). On the basis of the group discussions, the authors concluded that *"E-communications are not likely to be replacing the need to travel"*.

The third type of study in the literature that investigates the relationship between young adults' licence-acquisition and their online activity makes use of multivariate regression techniques (Table 1 summarises the features of recent studies employing multivariate techniques to investigate young adults' licence-holding). Using a sample of 15 countries, Sivak and Schoettle (2012) regress each country's rate of youth-licence-holding (expressed as a percentage) against the number of internet users in the country along with a set of other control variables. The authors report a negative all-else-equal cross-sectional association between internet-penetration and youth licencing, but the robustness of this study's experimental design and methodology have been the subject of vigorous debate (Le Vine et al. 2013, Sivak and Schoettle 2013a, Schoettle and Sivak 2013b).

In summary, there is a distinct lack of consensus as to whether online activity relates positively or negatively with driving licence holding. There is also confusion in this area of the scholarly literature regarding the distinction between cross-sectional statistical association and causality. For instance, the titles of (McDonald and Trowbridge, 2009) and (Sivak and Schoettle 2013b) respectively are *'Does the built environment affect when American teens become drivers?'* and *'The reasons for the recent decline in young driver licensing in the U.S.'*. Both imply causality whereas in each of these studies the authors show only cross-sectional association.

To the authors' knowledge the present study is the first in this body of literature to report findings from multi-variate regression analysis of disaggregate online-activity data.

3. Analysis of SHS dataset

This section describes the analysis undertaken to investigate the relationship between internet use and acquisition of a driving licence in Scotland. The dataset employed here is the 2005/06 version of the SHS (Hope, n.d.)

3.1 SHS data description

The SHS is a large-sample general social survey that has been undertaken continuously by the Scottish government since 1999. The SHS interview collects limited information about all members of interviewed households, and a higher level of detail for a single randomly-selected adult (age 16+) in each household.

The analysis draws from the "random-adult" SHS data for young adults between 17 and 29 years old. For the purposes of the analysis, the data has been weighted to represent the Scottish adult population in this age band. The unweighted sample size is 3,819 young adults.

The data employed in the analysis can be classified as three different types of information:

1. **Driving-licence-holding**: A binary variable of whether the respondent holds a full driving licence or not.
2. **Socio-demographic characteristics**. The set of covariates was chosen to match those employed in (Delbosc and Currie 2013) as closely as possible. This listing also includes a spatial-class variable (consistent with McDonald and Trowbridge [2009]) and highest-educational-qualification held:
    - Age
    - Gender
    - Household-level annual income, net of taxes and other deductions (converted into income quartiles)
    - Current economic status (full-time employment, part-time employment, secondary school, higher education, work-training scheme, unemployed, not in workforce),
    - Place of residence (large urban areas, other urban areas, small accessible rural towns, accessible rural areas, remote rural areas)

- Presence of ‘respondent’s children’ in household (proxied for via a binary variable that indicates whether each young-adult respondent lives in a household with any other member who is more than 15 years younger than them).
- Presence of ‘respondent’s parents’ in household (proxied for via a binary variable that indicates whether each young-adult respondent lives in a household with any other member who is more than 15 years older than them).
- Highest educational qualification ever obtained: ‘O’ grade or equivalent, ‘H’ grade / ‘A’ level or equivalent, Higher National Certificate/Higher National Diploma [HNC/HND], degree/professional-qualification/higher-degree.

3. **Internet-use characteristics.** Data collected here relates to personal use only, excluding online participation for working purposes. Questions of interest for the specific study are: (a) whether or not the random adult is an internet user, (b) the amount of time that he/she spends online per week, and (c) the participation in each of 17 specific classes of online activity (e.g. grocery-shopping, on-line learning, using e-mail etc.)

Table 2 contains descriptive statistics for the reported variables, including a cross-tabulation with licence-holding. Whilst space does not permit inclusion of a full correlation matrix, we note that the correlation coefficients between the control and internet-usage variables were modest in all cases, with the largest absolute value of any of the correlation coefficients being 0.32 (the positive correlation between current-economic-status-in-higher-education and use-of-internet-for-finding-information-related-to-education).

The most recent available SHS microdata are from 2008. However, for two reasons data from 2005/6 were employed in this study: (a) it is the most recent dataset available not affected by the recession beginning with the financial crash of 2007, and (b) the 2005/06 questionnaire module included the time-spent online variable, information that was not subsequently collected.

Though this dataset contains very rich detail of respondents’ online activity, it must be noted that much has changed in the online world since this dataset was collected in 2005/6. In the space of a small number of years very different regimes of online activities have emerged (e.g. video-calling and social networking) that are likely to interact in novel ways with physical mobility. Our results must be viewed with this in mind.

3.2. Logistic regression models of licence-holding using SHS dataset

In order to discriminate between aspects of licence-holding of internet users that are due to confounding socio-demographics and those linked *ceteris paribus* to online behaviour, a series of multivariate logistic regression analyses were performed. For all models, the dependent variable indicates whether each young-adult SHS respondent holds a full car driving licence or not. As with Delbosc/Currie (2013), all of the models are binary logistic form with cross-sectional parameter estimates and year-specific error terms.

Model run #1 contains only the socio-demographic and spatial control variables.

Model run #2 adds a single dummy parameter for internet usage into the model.

Model run #3 is identical to #2, with the addition of time-spent-online-per-week as a continuous quantity (with values allocated to the mid-points of each of the time-spent-online bands, and the 20-plus-hours-per-week category coded as 25-hours-per-week).

Model run #4 is identical to #3, but time-spent-online-per-week is entered as a series of dummy variables for each self-reported time band rather than a continuous quantity.

Model run #5 is identical to #4, with the addition of 17 dummy internet-activity variables for whether the respondent participated in the respective online activity or not.

Model runs #6 and #7 are analogous to #3 and #4 respectively, but using a sequential two-step estimation procedure. In the first step a linear regression model is estimated in which the internet-activity binary variables are used as independent variables to calculate a simulated value of time-spent-online-per-week for each respondent (the results of this first-step regression are shown in Table 4). In the second step a logistic regression model identical to model run #2 is estimated, but rather than respondents' self-reported time-spent-online-per-week includes the simulated calculation of this quantity. This two-step procedure is designed to address the potential for any cross-correlation between the time-spent and internet-activity parameters in models #6 and #7 leading to biased parameter estimates. (NB: the largest such correlation coefficient in the SHS dataset was 0.44, between frequency-of-internet-use and use-of-internet-for-emailing). In model run #6 time-spent-online is treated as a continuous quantity (as in model run #3). The range of predicted time-spent-online values proved not to be wide enough to estimate model run #7 using the same categories as model run #4, therefore in model run #7 time-spent-online is binned into four quartiles to investigate possible non-linearities with respect to time-spent-online.

The results of the baseline regression model (Model #1) are presented in Table 3. The rows associated with the internet usage parameters (of primary substantive interest in this analysis) are highlighted in **bold** text; control-variables are in un-bolded text. As with any cross-sectional logistic regression, it must be borne in mind that the direction of causality is not shown by the estimated parameters, which merely capture *ceteris paribus* statistical association.

The estimated effects of the socio-demographic control variables on licence-holding are consistent with those reported in Delbosc and Currie (2013), with two notable exceptions. The effects associated with living with 'one's children' and with 'one's parents' (NB: we employ proxy variables for these covariates, see the data description above) are both negative and statistically significant. By contrast, Delbosc/Currie (2013) found them both to be positive and statistically significant. It cannot be known whether these different effects are due to data differences or differences in contexts; the accumulation of further findings from other contexts will be necessary to reach a conclusion. Of the two control-variable effects that extend from the Delbosc/Currie (2013) set of covariates, with regards to the first (residence location) we find that living in larger settlements is nearly-monotonically negatively linked with licence-holding (the exception being the small-remote-towns spatial class), and that higher levels of educational attainment is also nearly-monotonically positively associated with licence-holding. The finding regarding spatial class is consistent with the findings of McDonald and Trowbridge (2009). With regards to the second such effect (highest level of educational attainment), we find a generally positive *ceteris paribus* relationship between educational attainment and licence-holding, with the largest difference being between holding a degree or higher and holding no educational qualification.

Model run #2 contains the simplest treatment of effects associated with internet usage. It was found that the estimated *ceteris paribus* statistical association between licence-holding and internet-usage is positive and statistically significant. A chi-squared test comparing models #1 and #2 also finds that model #2 is preferred to #1 ($p<0.05$). By converting the parameter estimate for internet-usage into an odds-ratio (cf. Hosmer et al. 2013), it can be interpreted as implying that the odds of an internet-using young adult holding a full-car-driving-licence are 70% larger than those of an otherwise identical young adult that does not use the internet.

Model run #3 brings in the self-reported information of how much time internet-using respondents spend online each week in the form of a continuous variable, along with a binary variable that distinguishes between internet-using and non-internet-using respondents. On the basis of a comparison of the adjusted pseudo-r^2 values (which balance between rewarding goodness-of-fit and penalising the addition of parameters), it can be concluded that model run #3 is statistically preferred to model run #2. As with model run #2, we find a positive, statistically significant effect associated with being an internet user. We also find a statistically-significant ($p<0.05$) negative relationship with the amount of time spent online per week. By taking the ratio of the two parameters (0.604/-0.012), it can be calculated that the combined effects are positive for internet users that spend less than 49 hours per week online and negative only for those that spend in excess of 49 hours per week online. This model specifies the marginal effect of time-spent-online to be linear; the next model investigates the possibility of non-linearities in this effect.

In model run #4 time-spent-online is entered as a set of independently-estimated parameters for each of the time-spent-online bands. All effects are relative to the reference category of non-internet-users. The adjusted pseudo-r^2 value of model run #4 is fractionally larger than that of model run #3 (0.2218 v. 0.2215), indicating that the additional variables provide enough statistical information that model run #4 is preferred to model run #3 in a statistical sense, but the difference is very small.

Two patterns in the results of model run #4 are noteworthy. First, the effects for all time-spent-online categories are positive, and all except the 20-plus-hours-per-week (the heaviest-usage category) parameter are statistically significant. Second, the relationship is an inverted 'U' shape; the effects increase monotonically from no time spent online (i.e. non-users), which is fixed at zero for normalisation, up to a peak for the effect associated with the five-to-ten-hours-per-week category. The effects then decrease monotonically for all remaining categories, with the effect of the heaviest time-online-per-week category (20-plus-hours) not being statistically distinguishable from zero.

Model run #5 extends from model run #4 by adding into the analysis the types of online activities that respondents indicate that they perform, in the form of 17 dummy variables. The adjusted pseudo-r^2 value indicates that this model structure is preferred to model run #4. We see that all of the effects associated with time-spent-online become statistically insignificant. The inverted 'U' shape of this relationship is also not strictly monotonic as it was in model run #4. Of the 17 online-activity dummy variables, four are found to be statistically significant at the $p<0.05$ level, two positively (personal-banking/financial-activities and finding-information-about-goods-and-services) and two negatively (playing/downloading-games and online-chat-rooms/sites). Two others are statistically significant at the $p<0.10$ level, both positively (e-mailing and buying/ordering-tickets-and-services). The remainder are not statistically significant, with p-values ranging from 0.15 to 0.85.

This model run (#5) includes a large set of freely-estimated parameters relating to internet-usage. Correlation between the online-activity variables and time-spent-online could lead to biased parameter estimates (empirically the largest of these correlation coefficients was found to be the correlation between time-spent-online and use-of-email, at 0.44). To address this possibility model run #6 employs a two-step estimation process, with the online-activity dummy variables used to calculate a prediction of time-spent-online for each internet-using respondent, and this variable included as a continuous variable in the second-step (the binary logistic model estimation), along with a separate binary indicator of whether or not a respondent uses the internet at all.

In the first estimation step (results shown in Table 4), 10 of the 17 online-activity parameters were found to be statistically-significantly (at $p<0.05$) associated with time-spent online (none were significant at $p<0.10$ but not $p<0.05$), with all but one positively-signed (the exception being looking-for-work-online). The strongest predictor (both in terms of magnitude, 3.2 hours/week, and statistical significance) was whether an internet user participates in online chat rooms/sites.

In the second estimation step, the results were similar to those of model run #3. Being an internet user was positively associated with licence-holding, whilst the marginal effect of time-spent-online was negative. The point at which the combined effect switches from positive to negative was estimated to be 20 hours of online activity per week, compared to 49 hours in model run #3. Model runs #3 and #6 have the same number of parameters, hence the larger pseudo-r^2 of model run #6 indicates that it is statistically preferred.

As with model run #6, model #7 is generated from a two-step estimation. The difference is that in model run #7 the predicted-time-spent-online variable is banded into quartiles (rather than continuous). Similar to model run #4 (which used self-reported time-spent-online categories), we see an inverted 'U' relationship whereby the estimated effect is largest for an intermediate category (the second quartile) and then decreases monotonically in both directions.

In order to illustrate the combined *ceteris paribus* relationship between licence-holding and these multiple aspects of internet-usage, a sample-enumeration procedure was performed. This was required because the net effects of internet usage arise from multiple parameters in model run #s 3 through #7, rather than a single parameter that applies in the same way to all internet users regardless of their internet-usage profile. The sample enumeration method involved multiplying the estimated parameters associated with internet-usage by the corresponding covariates representing the observed internet-usage profile for each young adult in the SHS sample, in order to estimate the idiosyncratic aggregate effect of internet usage for each respondent. For ease of interpretation, the effects were converted into odds-ratios (Hosmer et al. 2013). For each respondent, the odds ratio can be interpreted as the multiplicative effect of their idiosyncratic internet usage profile on the odds that they hold a driving licence. Values larger than 1.0 imply a positive *ceteris paribus* relationship between licence-holding and internet use, while the opposite applies to values below 1.0.

To expose this point via an example, let us consider the effect for a randomly chosen internet-using SHS respondent who is at the 30^{th} percentile of this distribution. This person is an employed 28-

year-old man, living in the 'other (non-large) urban areas' spatial class, living in a household with an annual household income of £28,000. He works full-time and lives with 'his children' but not 'his parents'. He holds a degree-or-professional-qualification-or-a-higher-degree. He reports using the internet less than an hour per week and the following types of online activities: finding information about goods and services, finding information related to education, general browsing or surfing, looking for work, using chat rooms or sites and using e-mails. This respondent is observed to hold a full driving licence and the full model specification predicts that with approximately 83% probability. Given his internet-usage profile, the model suggests that this respondent's odds of holding a full driving licence are 1.41 times higher than an otherwise identical respondent who does not use the internet.

The results of this sample-enumeration analysis are shown in Table 5. The aggregate odds ratio associated with internet usage is shown for ten percentile points on the distribution (10%, 20%, …, 80%, 90%). The 50th percentile is the median effect; it ranges between 1.50 (run #7) and 1.76 (run #3). Note that for model run #2, where there is a single parameter that applies to each internet user regardless of how they use the internet, the odds ratio is the same (1.70) across all internet users. It can be seen that the all-else-equal relationship between online activities and licence-holding is predicted to be positive for a large majority of internet users in the SHS sample, and this holds for all model runs using the SHS dataset. (NB: for all model runs but #5 the effect was strictly positive; for run #5 the odds ratio is less than 1.0 for the lowest 9% of the distribution).

These results from all model specifications with the SHS data suggest a positive cross-sectional relationship between online-activity and driving-licence holding, net of confounding effects.

4. Analysis of OPN dataset

This section outlines the results of a second set of analyses of the links between licence-holding and online activity, using the OPN data (which covers the entirety of Great Britain including England, Scotland and Wales,[ONS n.d.]).

4.1 OPN data description

It is plausible that the findings from the analysis of the SHS dataset might reflect the idiosyncratic characteristics of Scotland, or that they may be influenced by the survey instrument or the SHS' specific survey practices. As the findings from analysis of the SHS dataset are at odds with earlier results (Sivak and Schoettle 2013a), we sought to confirm the SHS results by reproducing as nearly as possible the same analysis with a second dataset (OPN) with different data-generation procedures and from a different temporal and spatial context.

A further distinction between the SHS and OPN datasets is that the OPN is newer (2008/9 v. 2005/6) and thus includes more recent developments in its definitions of online activity-classes (e.g. selling things online, video-conferencing, etc.).

The OPN dataset does not have the degree of richness the SHS contains in terms of travel information; unlike the SHS the instrument package does not include a travel diary. However a question on licence-holding is included, and can therefore be regressed against the detailed internet-usage behaviour the OPN captures.

The same age group (17 – 29 years old) was used to analyse both the OPN and SHS datasets and, as with the SHS, the OPN sample is appropriately weighted to be nationally-representative. Although the coverage of the OPN dataset is wider (all of Britain, which includes Scotland), the unweighted sample size of young adults is much smaller, approximately 1/4 of the SHS sample size (895 v. 3,819).

The type of data employed for the analysis is the same with the SHS with the following exceptions:

1. Income information is collected at the individual level rather than at the household level
2. Settlement size is not known. The only spatial information that is recorded is which of Britain's 11 Government Office Regions the respondent lives in. For model estimation, this variable was recoded into a binary variable of London/rest-of-Britain. Amongst other differences, London is served by a much more efficient public transport network than other parts of Britain.
3. The ages of other adult household members are not known, so therefore it was not possible to include a control variable for whether a young adult lives with their 'parents' (as defined in the SHS analysis).
4. The OPN dataset contains a single indicator of student status; we do not know (as in the SHS dataset) whether a student attends a secondary-school or higher-education. 'On work training scheme' is also not recorded as an economic status, as it is in the SHS.
5. The characterisation of highest-attained educational qualification is different; in the OPN data the following categories were recorded (in rough ascending order): No educational qualifications, GCSE grade D-G or CSE Grade 2-5 or Standard grade level 4-6, 'O' level or GCSE equivalent, ONC/National level BTEC, 'A' levels or higher, higher qualifications below degree level, degree level
6. The online-activity-classes (listed in Table 6) are not identical to the SHS online activity-classes; there are 31 classes rather than 17 in the SHS dataset., Also, the OPN survey did not include a question regarding time-spent-online.. Rather the most closely-analogous question is about frequency of online activity: "How often, on average, have you used the internet in the last 3 months".

Table 6 contains descriptive statistics for the relevant variables, including a cross-tabulation with licence-holding. The correlation coefficients between control and internet-usage variables were modest in all cases, with the largest absolute value of any of the correlation coefficients being 0.35 (for the positive correlation between highest-qualification-is-degree-or-higher and use-of-internet-for-purchasing-travel-or-holiday-accommodation).

4.2 Logistic regression model of licence-holding using OPN dataset

The following models were estimated (results are in Table 7); for ease of interpretation the same numbering scheme is applied to the SHS and OPN model runs:

Model run #1 is the baseline specification using the OPN dataset, with no indicators of internet usage. We note that the effect of living with one's 'children' was not found to be statistically significant; this is relevant as in the SHS dataset it was negative and statistically significant (and in the Delbosc/Currie [2013] analysis it was positive and statistically significant.

Model run #2 adds in a single binary indicator of whether each respondent self-reports having used the internet at least once in the past three months.

Model run #3 adds in a continuous variable of the number of days per week that each respondent reports using the internet. The continuous variable was constructed by taking the mid-point values of each of the frequency-of-use bands: Every day or almost every day (5.25 days/week), at least once a week but not every day (2.25 days/week), at least once a month but not every week (0.62 days/week), less than once a month but at least once in the last three months (0.16 days/week), not in last three months (0 days/week).

Model run #4 estimates separate effects for each of the frequency-of-internet-usage categories, to investigate possible non-linearities.

Model run #5 is identical to #4, with the addition of the binary online-activity-class indicators.

Model runs #6 and #7 employ the same two-step estimation procedure as model run #6 using the SHS dataset. This is to address the possibility for biased parameter estimates due to cross-correlation between online-activity-classes and frequency-of-internet use. (NB: as with the SHS dataset, the largest such correlation coefficient in the OPN dataset was between frequency-of-internet-use and use-of-internet-for-emailing, at +0.37). In model run #6 frequency-of-internet-use is treated as a continuous quantity, whilst in model run #7 it is classed into quartiles (as in the analysis of the SHS dataset).

The parameter estimates of the baseline model are in keeping with the SHS baseline model; none are estimated to be statistically identifiable from zero but with opposite signs.

In model run #2, the all-else-equal statistical relationship between internet-usage and licence-holding is estimated to be positive and statistically significant.

This is also the case in model run #3, and in this run the marginal effect of frequency-of-internet-use is also estimated to be positive but not statistically significant ($p=0.80$). This is somewhat different from the results of SHS model run #3, where the marginal effect of time-spent-online was found to be negative and statistically significant. It is worth noting that the adjusted pseudo-r^2 values of OPN runs #2 and #3 suggest that #2 is statistically preferred.

In model run #4 using the SHS data an inverted 'U' shape was found in the relationship between time-spent-online and licence-holding, and this was also found in run #4 using the OPN data with the frequency-of-internet-usage categories. The adjusted pseudo-r^2 values indicate that OPN model run #4 is statistically preferred to OPN runs #1, #2, and #3.

Model run #5 brings in the OPN dataset's binary variables for participation in each of the 31 online activity classes. The adjusted pseudo r^2 value of this run is the highest of all OPN runs, indicating that it is statistically preferred (though it may contain biased parameter estimates due to any correlation between the frequency-of-internet-use and participation in specific online activity classes.

Model run #6 addresses this possibility for bias using the two-step estimation process. Some information is sacrificed, as reflected in the smaller adjusted pseudo-r^2 values relative to run #s 3 and #4. What is found is a positive (but not statistically significant) effect associated with being an internet user, and a positive and statistically significant marginal effect associated with frequency-of-internet-usage.

In model run #7, the largest positive effect associated with frequency-of-internet-use is the highest quartile (those predicted in the first-step estimation to be the most frequent internet-users on the basis of the activities they perform online). This is different from the analogous SHS result, where the *smallest* effect was found to be associated with the highest quartile of time-spent-online. Table 8 contains the estimates of the first-step linear regression. It can be seen that the largest positive relationship is associated with using the internet for emailing (2.3 days of internet use/week).

Following the same sample-enumeration procedure as with the SHS analysis, the aggregate odds-ratios attributable to internet-usage are shown in Table 5. The table begins at the 10^{th} percentile of this distribution, and therefore it must be noted that for two model runs (#4 and #5) there are some internet users for whom the odds ratio is less than 1.0 (1% of users in run #4; 7% in run #5). For all models the median odds ratio is well in excess of 1.0, within the range between 3.15 (run #7) and 3.99 (run #6).

4.3 Comparison of estimated parameters from SHS and OPN models

The estimated *aggregate* internet usage effects with the OPN dataset are broadly consistent with the SHS results. It is, however, also of interest to investigate the relationships between the SHS and OPN results at the level of the individual internet-usage parameters relating to specific types of online activities.

It is not straightforward to compare all activity classes recorded in each dataset. First, the SHS records 17 activity classes against 31 in the OPN. Furthermore, not all of these 17 'SHS' activities match one-for-one with an OPN activity definition, even though in many cases they refer to similar types of activities.

Six activities were identified that matched one-for-one between the OPN and SHS datasets:

1) Sending and receiving e-mails
2) Looking for work
3) Looking for information related to education
4) Looking for information about goods or services
5) On-line learning
6) Grocery shopping.

Figure 1 is a scatterplot (with confidence intervals) showing the SHS and OPN estimated parameters for each of these six activity types.

When the value of the parameter is positively signed for both datasets the point will lie on the upper right quadrant of this scatterplot. Likewise, when the estimated values are both negative the point would be in the lower left quadrant.

Of the matching online activities, only one parameter has the same sign and is significant for both datasets: “sending and receiving e-mails” (NB: in the SHS analysis this parameter was significant at $p<0.10$ but not $p<0.05$). However, the line of best fit shown in Figure 1 through the six points has a positive slope, as would be expected. This implies that where they overlap the two sets of parameters are generally consistent with each other. The correlation between the two sets of parameters is 0.63, therefore the r^2 of the best fit line plotted in Figure 1 is 0.40. Interestingly, the slope of the best fit line is roughly one-sixth, meaning that the OPN parameters seem to be systematically larger in magnitude than the SHS parameters. In no case was an effect found to be positive and statistically-significant in one dataset and negative and statistically-significant in the other.

5. Conclusions

This study investigated the hypothesis of a negative cross-sectional statistical relationship between online activity and licence-holding among young adults.

We report analyses of two different nationally-representative databases, which both show a positive (or complementary) relationship between internet-usage and licence-holding, net of confounding effects. This finding is robust across a range of model specifications, including multi-stage estimations to address cross-correlation between indicators of internet usage. Both datasets are cross-sectional, and therefore it cannot be asserted that the cross-sectional relationship between online-activity and licence-holding identified in this study necessarily would hold were we able to observe individuals over time. In the latter case we would be able to more confidently assert a finding of causality. These findings however are the first to the authors’ knowledge to employ disaggregate multi-variate regression to investigate this research question.

In addition to this principal finding, several other noteworthy relationships were found. Only one online activity (using email) was found to be the same sign (positive) and statistically significant for both datasets (NB it was significant only at the $p<0.10$ level in the SHS dataset). No comparable effects due to any specific online activity were of opposite signs and statistically significant in the two datasets. Also, an inverted ‘U’ effect associated with time-spent-online was found using the SHS dataset, and with the OPN’s measure of internet-usage intensity (frequency of days/week using the internet) when the self-reported frequency-of-use covariate is entered directly into the estimation. But when a synthetic frequency-of-use variable is constructed using in a two-step estimation process (model run #7), this relationship did not hold in the OPN dataset. The accumulation of further evidence will be required to ascertain the generic nature of this relationship, which may well be context-dependent.

Our findings show a relationship different than the most comparable study (Sivak and Schoettle 2012) in the literature, though that study employed aggregate country-level analysis while our findings arise from analysis of disaggregate person-level microdata. These results require confirmation from other geographic contexts outside of Britain, and more recent data that take account of new developments in information technologies.

Acknowledgments

An earlier version of this material was presented at the 2014 Transportation Research Board Annual meeting (Le Vine et al. 2014). The authors wish to thank Transport Scotland, the RAC Foundation, the Independent Transport Commission, and the Office of Rail Regulation for financial support for this line of research, as well as four anonymous peer reviewers for helpful feedback. Any errors are the authors' sole responsibility.

About the authors

Scott Le Vine is a research associate in Imperial College London's Centre for Transport Studies, with research interests in long-run travel trends and shared-mobility systems. Further particulars can be found at: www.imperial.ac.uk/people/s.le-vine

Charilaos Latinopoulos is a doctoral candidate in Imperial College London's Centre for Transport Studies. His dissertation is addressing consumer strategies for charging electric-vehicles. Further particulars can be found at: www.imperial.ac.uk/people/charilaos.latinopoulos10

John Polak is Professor of Transport Demand, Chairman of the Centre for Transport Studies and Director of Research in the Department of Civil and Environmental Engineering at Imperial College London. He works in the areas of demand modelling, network control and traffic management and ITS.

References

Alcindor, Y. (2012) *Research shows that teens in no hurry to be behind the wheel.* Accessed 1 December 2013 via: http://usatoday30.usatoday.com/NEWS/usaedition/2012-03-15-Putting-off-driving_ST_U.htm

Beige, S., Axhausen, K.W. (2006) *Long-term mobility decisions during the life course: Experiences with a retrospective survey*. Paper presented at 11th International Conference on Travel Behaviour Research, Kyoto, 16-20 August.

Buccholz, T.G., and Buccholz, V. (2012) *The go-nowhere generation.* Accessed 1 December 2013 via: http://www.nytimes.com/2012/03/11/opinion/sunday/the-go-nowhere-generation.html?_r=0

Cao, X.J., Douma, F., Cleaveland, F., (2010) *The influence of E-shopping on Shopping Travel: Evidence from Twin Cities.* Transp Res Rec J Transp Res Board 2157, 147-154

Delbosc, A., and Currie, G. (2013a) *Causes of Youth Licensing Decline: A Synthesis of Evidence.* Transport Reviews. 33(3), p.271-290.

Delbosc, A., Currie, G. (2013b) *Changing demographics and young adult driver license decline in Melbourne, Australia (1994–2009).* Transportation. DOI: 10.1007/s11116-013-9496-z.

Delbosc, A., Currie, G. (2014) *Using discussion forums to explore attitudes towards cars and licencing among young Australians*. Transport Policy (in press)

Douma, F. (2003) *ICT and Travel in the Twin Cities Metropolitan Area: Enacted Patterns Between Internet Use and Working and Shopping Trips.* PhD Thesis. University of Minessota.

Grimsrud, M., El-Geneidy, A. (2013) Driving transit retention to renaissance: trends in Montreal commute public transport mode share and factors by age group and birth cohort. Public Transport. 5:3. p.219-241.

Hope, S (undated) *Scottish Household Survey: Methodology 2005/2006*. Prepared by Ipsos Mori for Scottish Executive.

Hosmer, D.W., Lemeshow, S., Sturdivant, R.X. (2013) *Applied Logistic Regression*, 3rd edition, Wiley. ISBN: 978-0-470-58247-3. Hoboken, New Jersey.

Institute for Mobility Research (ifmo) 2013. *'Mobility Y' – The Emerging Travel Patterns of Generation Y. Munich*

Kay, A.I., Noland, R.B., Rodier, C.J. (2014) *Achieving reductions in greenhouse gases in the US road transportation sector.* Energy Policy. DOI: 10.1016/j.enpol.2014.02.012.

Kuhnimhof et al (2012). *Men Shape a Downward Trend in Car Use among Young Adults – Evidence from Six Industrialized Countries,* Transport Reviews, 32 (6), 761-779.

Latinopoulos, C., Le Vine, S., Jones, P., Polak, J. (2013) *Car, rail and bus travel trends in Scotland, 1995 to 2010.* Prepared for Transport Scotland.

Le Vine, S., Jones, P. (2012). *On the Move: Making sense of car and train travel trends in Britain.* Accessed 1 December 2013 via: http://www.racfoundation.org/research/mobility/on-the-move-main-research-page

Le Vine, S., Latinopoulos, C., Polak, J. (2013). *A tenuous result: re-analysis of the link between internet-usage and young adults' driving-licence holding. Comments on 'Recent changes in the age composition of drivers in 15 countries'*. Traffic Injury Prevention.

Le Vine, S., Latinopoulos, C., Polak, J. (2013). *What is the relationship between online activity and driving-licence-holding amongst young adults?*. Paper presented at the 93rd Annual meeting of the Transportation Research Board, Washington, DC. January 2014.

Le Vine, S., Lee-Gosselin, M., Sivakumar, A., Polak, J. (2013) *A new concept of accessibility to personal activities: development of theory and application to an empirical study of mobility resource holdings,* Journal of Transport Geography, Volume 31, July 2013, 1-10.

Litman, T. (2012) *Current mobility trends: Implications for Sustainability.* In: Keep moving, towards sustainable mobility, edited by Bert van Wee.

McDonald, N. And Trowbridge, M. (2009) *Does the built environment affect when American teens become drivers?* Evidence from the 2001 National Household Travel Survey. Journal of Safety Research, 40:177-183.

Moktharian, P. (2002) *Telecommunications and Travel: The Case for Complementarity.* Industrial Ecology, 6 (2) p.43-57.

Mokhtarian, P. (1990) *A typology of relationships between telecommunications and transportation.* Transportation Research Part A, 24A(3), p. 231-242.

Office for National Statistics [ONS] (undated) *National Statistics Opinions Survey - Technical Report January 2009*.

Puentes, R. (2012) *Have Americans hit peak travel? A discussion of the changes in US driving habits.* OECD International Transport Forum Discussion Paper 2012-14. Prepared for the Roundtable on Long-Run Trends in Travel Demand, 29-30 Nov 2012.

Sivak, M., Schoettle, B. (2012) *Recent changes in the age composition of drivers in 15 countries.* Traffic Injury Prevention. 12:125-132.

Sivak, M., Schoettle, B. (2013a) *The Reasons for the Recent Decline in Young Driver Licensing in the U.S.* Traffic Injury Prevention, 15:1, p.6-9.

Sivak, M., Schoettle, B. (2013b *Towards understanding the recent large reductions in the proportion of young persons with a driver's license: A response to Le Vine et al. (2013)*. Trafic Injury Prevention, 14:6, p.658-659.

Tefft, B., Williams, A.F., Grabowski, J.G. (2013) Timing of Driver's License Acquisition and Reasons for Delay among Young People in the United States, 2012. Prepared for AAA Foundation for Traffic Safety. Accessed 1 December 2013 via: www.aaafoundation.org/sites/default/files/Teen%20Licensing%20Survey%20FINAL_0.pdf

Tefft, B., Williams, A.F., Grabowski, J.G. (2014) *Driver licensing and reasons for delaying licensure among young adults ages 18-20, United States, 2012*. Injury Epidemiology, 1:4. DOI: 10.1186/2197-1714-1-4

Vovsha, P., Petersen, E. (2009) *Model for person and household mobility attributes.* Transportation Research Record: Journal of the Transportation Research Board, 2132(1), 95-105

Williams, A.F. (2011) *Teenagers' Licensing Decisions and Their Views of Licensing Policies: A National Survey.* Traffic Injury Prevention. 12:4. p.312-319.

Tables and Figures

Reference	Study title	Data	Multivariate statistical technique
McDonald and Trowbridge (2009)	*Does the built environment affect when American teens become drivers?*	Person-level data, US National Household Travel Survey, 2001 (n=3,976)	Cross-sectional propensity-scoring and direct-matching of respondents with observably-similar demographic profiles living in different types of built environments
Sivak and Schoettle (2012)	*Recent changes in the age composition of drivers in 15 countries*	Country-level data of varying years (circa 2010), combined from various sources (n=15)	Cross-sectional stepwise-with-backwards-elimination linear regression
Tefft et al. (2013)	*Timing of Driver's License Acquisition and Reasons for Delay among Young People in the United States, 2012*	Person-level data, bespoke survey, 2012 (n=1,039)	Cross-sectional binary logistic regression
Delbosc and Currie (2013b)	*Changing demographics and young adult driver license decline in Melbourne, Australia (1994–2009)*	Person-level data, Victorian Area Travel Survey / Victorian Integrated Survey of Travel and Activity, 1994/1999/2007/2009 (n=14,818)	Cross-sectional binary logistic regression with year-specific error terms
This study	***What is the relationship between online activity and driving-licence-holding amongst young adults?***	**Person-level data, Scottish Household Survey, 2005/6 (n=3,819) and Opinions and Lifestyle Survey, 2008/9 (n=895)**	**Cross-sectional binary logistic regression with year-specific error terms**

Table 1: Summary of recent studies of young adults' licence-holding which employed multivariate statistical techniques

Indicator	All young adults in sample	Holders of a full car driving licence	Non-holders of a full car driving licence
Percentage that use the internet	73%(1%)	82%(1%)	64%(1%)
Age	23.1 (0.06)	24.2 (0.08)	21.9 (0.09)
Income	£22,344 (£232)	£25,024 (£318)	£19,389 (£324)
Percentage that hold a full car driving licence	52% (1%)	-	-
Percentage female	53%(1%)	51% (1%)	56% (1%)
Percentage living in large urban areas	48% (1%)	45% (1%)	51% (1%)
Percentage living in other urban areas	28% (1%)	27% (1%)	29% (1%)
Percentage living in small accessible towns	8% (<0.5%)	9% (1%)	7% (1%)
Percentage living in small remote towns	4% (<0.5%)	3% (<0.5%)	5% (1%)
Percentage living in accessible rural areas	8% (<0.5%)	11% (1%)	6% (1%)
Percentage living in remote rural areas	4% (<0.5%)	5% (1%)	2% (<0.5%)
Percentage that are employed full-time	50% (1%)	66% (1%)	32% (1%)
Percentage that are employed part-time	9% (1%)	8% (1%)	10% (1%)
Percentage that are not in workforce	11% (1%)	6% (1%)	16% (1%)
Percentage that are unemployed	8% (<0.5%)	3% (<0.5%)	13% (1%)
Percentage that are students at secondary school	3% (<0.5%)	0% (<0.5%)	6% (1%)
Percentage that are students in higher education	19% (1%)	17% (1%)	21% (1%)
Percentage that are in government work/training scheme	0% (<0.5%)	0% (<0.5%)	1% (<0.5%)
Percentage living with their 'own children'	23% (1%)	20% (1%)	26% (1%)
Percentage living with 'parents'	41%(1%)	35%(1%)	48%(1%)
Percentage that hold 'O' grade or equivalent qualification	25% (1%)	18% (1%)	34% (1%)
Percentage that hold H' grade/ 'A' level or equivalent qual.	31% (1%)	32% (1%)	30% (1%)
Percentage that hold HNC/HND or equivalent qual.	13% (1%)	16% (1%)	9% (1%)
Percentage that hold Degree or professional. qual. or higher deg.	20% (1%)	28% (1%)	10% (1%)
Percentage that hold no qualification	11% (1%)	6% (1%)	17% (1%)
Travel behaviour (from SHS travel diary instrument)			
Average annual car driving mileage	1,808 (77.6)	3,529 (139.3)	41 (17.5)
Average annual number of car driving journeys	302 (9.9)	591(16.8)	7 (1.9)
Average annual mileage (all but car driving)	1,964 (79.1)	1,493 (179.8)	2,447 (110.8)
Average annual number of journeys (all but car driving)	507 (12.1)	329 (14.8)	691 (18.3)

Classes of online-activity			
Buying or ordering tickets and services	46% (1%)	58% (1%)	33% (1%)
Finding information about goods/services	47% (1%)	57% (1%)	36% (1%)
Finding information related to education	39% (1%)	43% (1%)	35% (1%)
General browsing or surfing	59% (1%)	68% (1%)	49% (1%)
Grocery shopping	10% (1%)	13% (1%)	7% (1%)
Looking for work	35% (1%)	42% (1%)	28% (1%)
Non-grocery shopping	31% (1%)	40% (1%)	22% (1%)
Online learning	18% (1%)	19% (1%)	16% (1%)
Paying rent	2% (<0.5%)	2% (<0.5%)	2% (<0.5%)
Personal banking / financial / investment activities	28% (1%)	39% (1%)	17% (1%)
Playing or downloading games	24% (1%)	24% (1%)	25% (1%)
Playing or downloading music	42% (1%)	45% (1%)	40% (1%)
Using chat rooms or sites	21% (1%)	18% (1%)	24% (1%)
Using e-mail	65% (1%)	74% (1%)	54% (1%)
Using or accessing government/official sites	21% (1%)	27% (1%)	15% (1%)
Voting	1% (<0.5%)	1% (<0.5%)	1% (<0.5%)
None of these	1% (<0.5%)	1% (<0.5%)	1% (<0.5%)
Time spent online			
Up to 1 hour per week	15% (1%)	16% (1%)	13% (1%)
Over 1 hour, up to 5 hours per week	28% (1%)	33% (1%)	23% (1%)
Over 5 hours, up to 10 hours per week	15% (1%)	18% (1%)	12% (1%)
Over 10 hours, up to 20 hours per week	9% (<0.5%)	10% (1%)	8% (1%)
Over 20 hours per week	6% (<0.5%)	5% (1%)	7% (1%)

Table 2: Descriptive statistics of SHS sample of Scottish young adults (age 17 to 29). Values in brackets are standard errors

Model run #	#1	#2	#3	#4	#5	#6	#7
Null log-likelihood				-2,914.93 for all model runs			
Final log-likelihood	-2,262.83	-2,246.15	-2,246.26	-2,240.35	-2.205.77	-2,242.94	-2,239.74
McFadden's pseudo r^2	0.2237	0.2294	0.2301	0.2314	0.2433	0.2305	0.2316
McFadden's adjusted pseudo r^2	0.2158	0.2212	0.2215	0.2218	0.2278	0.2220	0.2224

Parameter estimates							
Constant	-3.044**	-3.317**	-3.345	-3.323	-2.881	-3.284	-3.302
Lowest income quartile	Reference						
Second income quartile	0.108	0.094	0.098	0.101	0.071	0.094	0.099
Third income quartile	0.506**	0.453**	0.453**	0.449**	0.402**	0.458**	0.464
Fourth income quartile	0.860**	0.772**	0.775**	0.776**	0.731**	0.780**	0.784
Employed full-time	Reference						
Employed part-time	-0.675**	-0.686**	-0.684**	-0.687**	-0.655**	-0.678**	-0.682
Not in workforce	-1.367**	-1.332**	-1.321**	-1.325**	-1.285**	-1.331**	-1.331**
Unemployed	-1.677**	-1.658**	-1.630**	-1.606**	-1.545**	-1.648**	-1.652**
Secondary school student	-2.209**	-2.301**	-2.274**	-2.307**	-2.108**	-2.255**	-2.247**
Higher education student	-0.358**	-0.427**	-0.389**	-0.405**	-0.353**	-0.392**	-0.393**
On work training scheme	-1.881**	-1.947**	-1.953**	-1.959**	-1.825**	-1.933**	-1.950**
Lives with 'own children'	-0.236**	-0.176	-0.176	-0.171	-0.105	-0.176	-0.169
Lives with 'parents'	-0.225**	-0.226**	-0.218**	-0.230**	-0.187**	-0.228**	-0.219**
Female	-0.102	-0.119	-0.140*	-0.141*	-0.214**	-0.149*	-0.158
Age	0.159**	0.161**	0.162**	0.161**	0.144**	0.159**	0.159**
Year 2005	Reference						
Year 2006	0.054	0.033	0.044	0.032	0.025	0.044	0.042
Residence in remote rural areas	Reference						
Accessible rural areas	-0.217	-0.196	-0.197	-0.210	-0.205	-0.189	-0.179
Remote small towns	-1.092**	-1.122**	-1.106**	-1.124**	-1.049**	-1.108**	-1.093**
Accessible small towns	-0.555**	-0.530**	-0.521**	-0.523**	-0.507**	-0.508**	-0.485**
Other (not large) urban areas	-0.955**	-0.971**	-0.960**	-0.966**	-0.962**	-0.953**	-0.931**
Large urban areas	-0.989**	-0.977**	-0.962**	-0.975**	-0.968**	-0.950**	-0.921**
No educational qualifications	Reference						
'O' grade or equivalent	0.085	0.008**	0.009	0.019	0.008	0.015	0.010
'H' grade / 'A' level or equivalent	0.828**	0.693**	0.697**	0.704**	0.633**	0.715**	0.708**
HNC/HND or equivalent	0.789**	0.621**	0.623**	0.621**	0.601**	0.650**	0.642**
Degree/professional qualification/higher degree	1.089**	0.897**	0.908**	0.900**	0.747**	0.928**	0.910**
Status as internet-user	--	0.529**	0.604**	--	--	0.762**	--
Time spent online per week (hours/week)	--	--	-0.012**	--	--	--	--
Time spent online per week (up to 1 hour)	--	--	--	0.525**	0.123	--	--
Time spent online per week (over 1 hour, up to 5 hours)	--	--	--	0.526**	0.036	--	--
Time spent online per week (over 5 hours, up to 10)	--	--	--	0.709**	0.238	--	--
Time spent online per week (over 10 hours, up to 20)	--	--	--	0.531**	0.106	--	--
Time spent online per week (over 20 hours)	--	--	--	0.101	-0.260	--	--
Predicted time spent online per week (hours/week)	--	--	--	--	--	-0.038**	--
Predicted time spent online per week (lowest quartile)	--	--	--	--	--	--	0.635**

Predicted time spent online per week (second quartile)	--	--	--	--	--	--	0.720**
Predicted time spent online per week (third quartile)	--	--	--	--	--	--	0.408**
Predicted time spent online per week (highest quartile)	--	--	--	--	--	--	0.341**
Buying or ordering tickets	--	--	--	--	0.184*	--	--
Finding information about goods and services	--	--	--	--	0.202**	--	--
Finding information related to education	--	--	--	--	-0.153	--	--
General browsing or surfing	--	--	--	--	0.168	--	--
Grocery shopping	--	--	--	--	-0.189	--	--
Looking for work	--	--	--	--	0.055	--	--
Non-Grocery shopping	--	--	--	--	0.037	--	--
Online learning	--	--	--	--	0.093	--	--
Paying rent	--	--	--	--	-0.181	--	--
Personal banking/financial/investment activities	--	--	--	--	0.486**	--	--
Playing or downloading games	--	--	--	--	-0.286**	--	--
Playing or downloading music	--	--	--	--	-0.106	--	--
Using chat rooms or sites	--	--	--	--	-0.321**	--	--
Using email	--	--	--	--	0.268*	--	--
Using or accessing governmental/official sites	--	--	--	--	-0.119	--	--
Voting	--	--	--	--	0.166	--	--
None of these	--	--	--	--	0.726	--	--

Table 3: Results from logistic regression models of full-car-driving–licence-holding by Scottish young adults (age 17-29), SHS data. Values marked with * or ** are statistically significant at p<0.10 and p<0.05 respectively.

r^2	0.1799
Adjusted r^2	0.1754
Parameter estimates	
Constant	2.044
Buying or ordering tickets	-0.016
Finding information about goods and services	0.408

Finding information related to education	0.799**
General browsing or surfing	0.144
Grocery shopping	0.279
Looking for work	-0.539**
Non-Grocery shopping	0.642**
Online learning	1.282**
Paying rent	2.017**
Personal banking/financial/investment activities	0.565**
Playing or downloading games	2.332**
Playing or downloading music	1.067**
Using chat rooms or sites	3.181**
Using email	1.002**
Using or accessing governmental/official sites	-0.146
Voting	0.320
None of these	-0.448

Table 4: Results from linear regression model of SHS-dataset internet-users' time-spent-online, using the online activities in which they participate as independent variables. Values marked with * or ** are statistically significant at p<0.10 and p<0.05 respectively.

SHS	**10^{th} percentile**	**20^{th} percentile**	**30^{th} percentile**	**40^{th} percentile**	**50^{th} percentile**	**60^{th} percentile**	**70^{th} percentile**	**80^{th} percentile**	**90^{th} percentile**
Model run #1	--	--	--	--	--	--	--	--	--
Model run #2	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70
Model run #3	1.53	1.53	1.67	1.67	1.76	1.76	1.76	1.76	1.82
Model run #4	1.69	1.69	1.69	1.69	1.69	1.69	1.70	2.03	2.03

Model run #5	0.99	1.23	1.41	1.57	1.75	1.96	2.21	2.56	3.16
Model run #6	1.41	1.51	1.58	1.64	1.70	1.76	1.81	1.85	1.90
Model run #7	1.41	1.41	1.50	1.50	1.50	1.89	1.89	2.05	2.05
OPN									
Model run #1	--	--	--	--	--	--	--	--	--
Model run #2	3.25	3.25	3.25	3.25	3.25	3.25	3.25	3.25	3.25
Model run #3	3.16	3.32	3.32	3.32	3.32	3.32	3.32	3.32	3.32
Model run #4	3.25	3.25	3.25	3.25	3.25	3.25	3.25	3.25	4.74
Model run #5	1.12	1.65	2.23	2.96	3.82	4.94	6.46	9.23	18.41
Model run #6	3.18	3.49	3.68	3.82	3.99	4.14	4.28	4.47	4.76
Model run #7	2.71	2.71	3.02	3.02	3.15	3.15	3.15	4.05	4.05

Table 5: Amongst internet-users, percentile distribution of *ceteris paribus* relationship between internet usage and licence-holding, characterised as odds ratios, for all SHS and OPN model runs

Indicator	All young people in the sample	Holders of a full driving licence	Non-holders of a full driving licence
Percentage that use the internet	93% (1%)	97% (1%)	87% (2%)
Age	23.0 (0.12)	23.8 (0.15)	22.0 (0.20)
Income	£13,598 (£367)	£16,153 (£490)	£9,480 (£446)

Percentage that hold a full car driving licence	58% (2%)	-	-
Percentage that are female	50% (2%)	47% (2%)	54% (3%)
Percentage that reside in London	16% (1%)	14% (2%)	19% (2%)
Percentage that are employed full-time	58% (2%)	71% (2%)	40% (3%)
Percentage that are employed part-time	23% (1%)	20% (2%)	27% (2%)
Percentage that are not in workforce	11% (1%)	6% (1%)	18% (2%)
Percentage that are unemployed	3% (1%)	1% (<0.5%)	6% (1%)
Percentage that are full time students	5% (1%)	2% (1%)	9% (2%)
Percentage living with their 'own children'	18% (1%)	17% (2%)	19% (2%)
Percentage that have degree level qualification (or equivalent)	22% (1%)	28% (2%)	13% (2%)
Percentage that have higher educational qualification below degree level	7% (1%)	9% (1%)	5% (1%)
Percentage that have A-Levels or Higher	22% (1%)	23% (2%)	20% (2%)
Percentage that have ONC/National level BTEC	5% (1%)	6% (1%)	3% (1%)
Percentage that have O level or GCSE equivalent (Grade A-C) or O Grade/CSE equivalent	21% (1%)	17% (2%)	26% (2%)
Percentage that have GCSE grade D-G or CSE grade 2-5 or Standard Grade level 4-6	6% (1%)	6% (1%)	5% (1%)
Percentage that have other qualifications	8% (1%)	6% (1%)	11% (2%)
Percentage that have no formal qualifications	10% (1%)	5% (1%)	17% (2%)
Class of online-activity			
Sending and receiving e-mails	85% (1%)	92% (1%)	75% (2%)
Finding information about goods or services	71% (2%)	79% (2%)	60% (3%)
Using services related to travel and accommodation	56% (2%)	65% (2%)	44% (3%)
Downloading software (no games)	44% (2%)	50% (2%)	36% (3%)
Reading or downloading online news	51% (2%)	55% (2%)	45% (3%)
Looking for job or sending application	41% (2%)	45% (2%)	35% (3%)
Seeking health related information	31% (2%)	37% (2%)	23% (3%)
Looking for information about education	42% (2%)	42% (2%)	42% (1%)
Doing an online course	9% (1%)	9% (1%)	8% (2%)
Consulting the internet for learning	37% (2%)	40% (2%)	32% (3%)
Internet Banking	51% (2%)	61% (2%)	36% (2%)
Selling goods or services	19% (1%)	25% (2%)	12% (2%)
Obtaining information from public authorities	36% (2%)	43% (2%)	27% (2%)
Downloading official forms	24% (1%)	30% (2%)	15% (2%)
Sending in filled forms	19% (1%)	22% (2%)	14% (2%)
Telephoning and video calls	24% (1%)	25% (2%)	23% (3%)
Uploading self-created content	47% (2%)	53% (2%)	39% (3%)
Listening to web radios/watching web tv	46% (2%)	50% (2%)	41% (3%)
Playing or downloading games/images/films/music	63% (2%)	66% (2%)	58% (2%)

Foods or groceries	11% (1%)	12% (1%)	11% (2%)
Household goods	24% (1%)	31% (2%)	14% (2%)
Films/music	40% (2%)	48% (2%)	27% (2%)
Books/magazines/newspapers/e-learning	25% (1%)	31% (2%)	18% (2%)
Clothes/sports goods	39% (2%)	45% (2%)	32% (2%)
Computers, software and upgrades	24% (1%)	27% (2%)	20% (1%)
Computer hardware	10% (1%)	11% (1%)	8% (2%)
Electronic equipment	21% (1%)	25% (2%)	15% (2%)
Share purchases/financial services/insurance	9% (1%)	14% (2%)	1% (2%)
Travel or holiday accommodation	31% (2%)	38% (2%)	20% (2%)
Tickets for events	27% (2%)	34% (2%)	18% (2%)
Lotteries or betting	11% (1%)	15% (2%)	6% (1%)
Frequency of internet use			
Every day, or almost every day	75% (1%)	78% (2%)	70% (2%)
At least once a week, but not every day	15% (1%)	16% (2%)	12% (2%)
At least once a month, but not every week	2% (1%)	2% (1%)	3% (1%)
Less than once a month, at least once in past three months	1% (<0.5%)	1% (<0.5%)	2% (1%)

Table 6: Descriptive statistics of OPN sample of British young adults (age 17 to 29). Values in brackets are standard errors

Model run #	#1	#2	#3	#4	#5	#6	#7
Null log-likelihood	-482.80 for all model runs						
Final log-likelihood	-396.39	-391.32	-391.29	-386.63	-343.17	-388.78	-390.58

McFadden's pseudo r^2	0.1790	0.1895	0.1895	0.1992	0.2892	0.1947	0.1910
McFadden's adjusted pseudo r^2	0.1396	0.1481	0.1460	0.1516	0.1774	0.1512	0.1434
Parameter estimates							
Constant	-2.171	-3.280	-3.296	-3.412	-3.485	-3.524	-3.301
Lowest income quartile	Reference						
Second income quartile	0.297	0.262	0.260	0.272	0.272	0.274	0.267
Third income quartile	0.757**	0.655**	0.655**	0.621**	0.538**	0.642**	0.659**
Fourth income quartile	2.010**	1.890**	1.884**	1.875**	2.236**	1.831**	1.867**
Employed full-time	Reference						
Employed part-time	-0.383*	-0.457**	-0.459**	-0.473**	-0.548**	-0.469**	-0.463**
Full time student	-1.863**	-1.923**	-1.924**	-1.955**	-1.986*	-1.985**	-1.970**
Unemployed	-1.563*	-1.471	-1.472	-1.316	-1.009	-1.325	-1.472
Not in workforce	-0.596*	-0.612*	-0.613*	-0.639*	-0.624	-0.624*	-0.652*
Lives with 'own children'	-0.049	0.043	0.051	0.120	0.149	0.136	0.056
Female	-0.033	-0.018	-0.016	-0.025	0.016	0.015	0.015
Age	0.069**	0.080**	0.080**	0.082**	0.087**	0.089**	0.083**
Year 2008	Reference						
Year 2009	-0.020	-0.075	-0.083	-0.071	-0.152	-0.134	-0.103
All of Britain except London	Reference						
Residence in London	-1.172**	-1.208**	-1.208**	-1.207**	-1.362**	-1.226**	-1.250**
No educational qualifications	Reference						
Degree level	1.031**	0.823**	0.808**	0.886**	0.504	0.615*	0.744
Higher qualifications (below degree level)	1.061**	0.862**	0.859**	0.946**	0.661**	0.736*	0.816
'A' levels or higher	1.487**	1.275**	1.264**	1.340**	1.275**	1.114**	1.242**
ONC / National level BTEC	1.769**	1.565**	1.563**	1.723**	2.005	1.506**	1.572**
'O' level or GCSE equivalent	0.431	0.291	0.289	0.354	0.341	0.263	0.279
GCSE Grade D-G or CSE grade 2-5 or Standard grade level 4-6	1.345**	1.205**	1.209**	1.207**	1.321**	1.230**	1.219**
Other qualifications	0.094	-0.018	-0.024	0.057	-0.053	-0.049	-0.026
Internet-user status		1.179**	1.111**	--	--	0.332	--
Frequency of internet use (days of internet use per week)			0.017	--	--	--	--
Frequency of internet use (less than once a month, at least once in last 3 months)	--	--	--	-0.588	-0.525	--	--
Frequency of internet use (at least once a month, but not every week)	--	--	--	0.344	0.042	--	--

Frequency of internet use (at least once a week, but not every day)	--	--	--	1.556**	1.291**	--	--
Frequency of internet use (every day, or almost every day)	--	--	--	1.180**	0.712	--	--
Predicted frequency of internet use (days/week)	--	--	--	--	--	0.224**	--
Predicted frequency of internet use (lowest quartile)	--	--	--	--	--	--	1.104**
Predicted frequency of internet use (second quartile)	--	--	--	--	--	--	1.147**
Predicted frequency of internet use (third quartile)	--	--	--	--	--	--	0.998**
Predicted frequency of internet use (highest quartile)	--	--	--	--	--	--	1.398**
Sending and receiving e-mails	--	--	--	--	0.796**	--	--
Finding information about goods or services	--	--	--	--	-0.246	--	--
Using services related to travel and accommodation	--	--	--	--	-0.327	--	--
Downloading software (no games)	--	--	--	--	0.889**	--	--
Reading or downloading online news	--	--	--	--	-0.436*	--	--
Looking for job or sending application	--	--	--	--	0.052	--	--
Seeking health related information	--	--	--	--	0.294	--	--
Looking for information about education	--	--	--	--	-0.188	--	--
Doing an online course	--	--	--	--	-0.889**	--	--
Consulting the internet for learning	--	--	--	--	0.252	--	--
Internet Banking	--	--	--	--	0.444*	--	--
Selling goods or services	--	--	--	--	0.234	--	--
Obtaining information from public authorities	--	--	--	--	0.011	--	--
Downloading official forms	--	--	--	--	0.483*	--	--
Sending in filled forms	--	--	--	--	0.012	--	--
Telephoning and video calls	--	--	--	--	-0.175	--	--
Uploading self-created content	--	--	--	--	-0.048	--	--
Listening to web radios/watching web tv	--	--	--	--	-0.351	--	--
Playing or downloading games/images/films/music	--	--	--	--	-0.474	--	--
Foods or groceries	--	--	--	--	-1.156**	--	--
Household goods	--	--	--	--	0.716**	--	--
Films/music	--	--	--	--	0.280	--	--
Books/magazines/newspapers/e-learning	--	--	--	--	0.445*	--	--
Clothes/sports goods	--	--	--	--	-0.299	--	--
Computers, software and upgrades	--	--	--	--	-0.551**	--	--
Computer hardware	--	--	--	--	-0.344	--	--
Electronic equipment	--	--	--	--	0.227	--	--
Share purchases/financial services/insurance	--	--	--	--	1.995**	--	--
Travel or holiday accommodation	--	--	--	--	-0.259	--	--

Tickets for events	--	--	--	--	0.106	--	--
Lotteries or betting	--	--	--	--	0.405	--	--

Table 7: Results from logistic regression models of full-car-driving–licence-holding by British young adults (age 17-29), OPN data. Values marked with * or ** are statistically significant at p<0.10 and p<0.05 respectively.

r^2	0.5069
Adjusted r^2	0.4892

Parameter estimates	
Constant	1.275**
Sending and receiving e-mails	2.304**
Finding information about goods or services	0.068
Using services related to travel and accommodation	0.284**
Downloading software (no games)	0.058
Reading or downloading online news	0.417**
Looking for job or sending application	0.094
Seeking health related information	-0.023
Looking for information about education	0.153
Doing an online course	0.243
Consulting the internet for learning	0.200*
Internet Banking	0.051
Selling goods or services	0.043
Obtaining information from public authorities	0.202*
Downloading official forms	0.031
Sending in filled forms	-0.043
Telephoning and video calls	0.049
Uploading self-created content	0.051
Listening to web radios/watching web tv	0.030
Playing or downloading games/images/films/music	0.307*
Foods or groceries	0.119
Household goods	-0.003
Films/music	-0.003
Books/magazines/newspapers/e-learning	-0.075
Clothes/sports goods	0.111
Computers, software and upgrades	0.189
Computer hardware	-0.122
Electronic equipment	0.130
Share purchases/financial services/insurance	-0.136
Travel or holiday accommodation	-0.121
Tickets for events	-0.016
Lotteries or betting	0.026

Table 8: Results from linear regression model of OPN-dataset internet-users' frequency of internet use, using the online activities in which they participate as independent variables. Values marked with * or ** are statistically significant at p<0.10 and p<0.05 respectively.

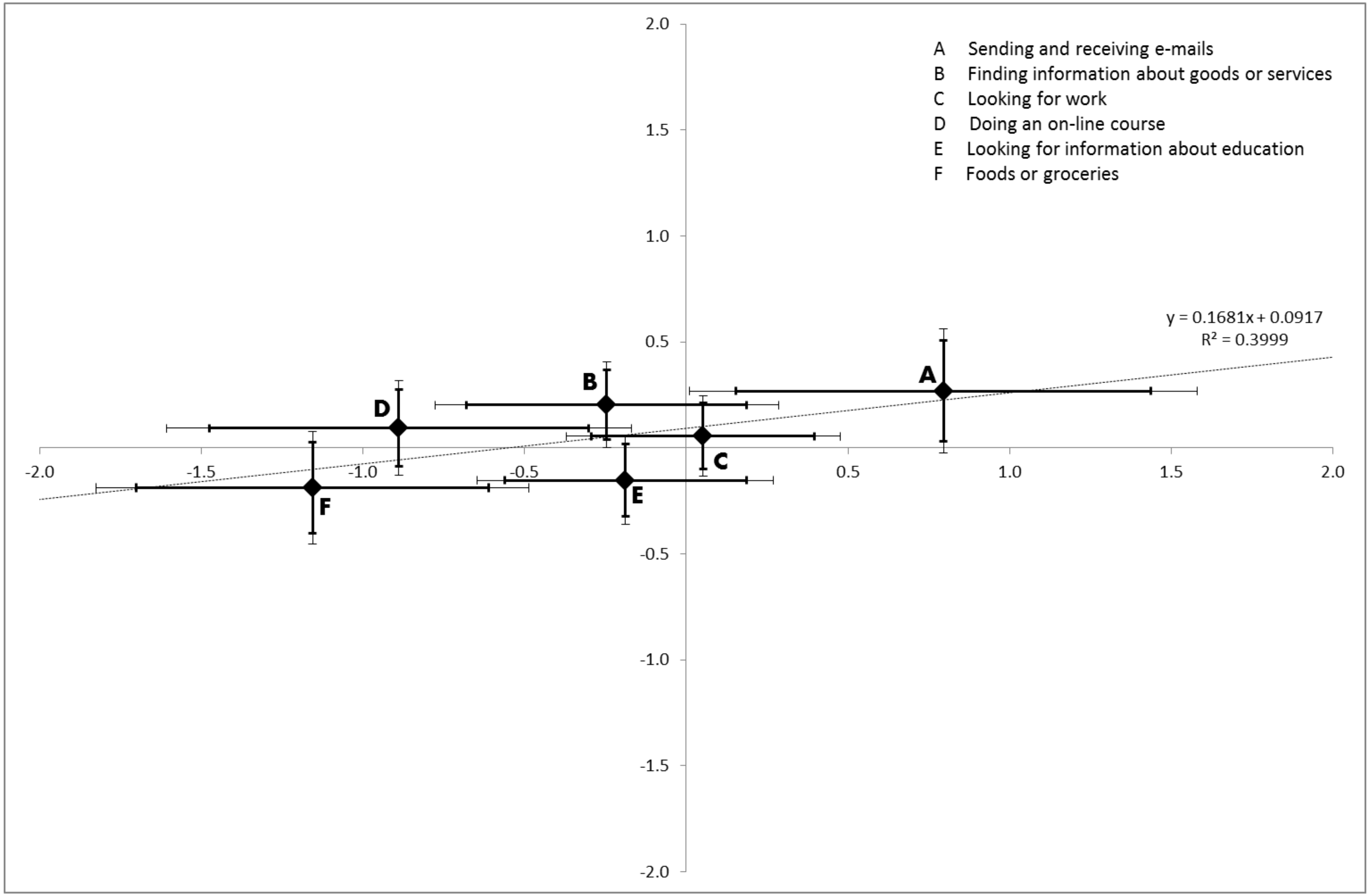

Figure 1: Scatterplot of parameter estimates for the six classes of online activities where a one-to-one correspondence exists between the SHS and OPN datasets (parameter estimates from model run #5 of both datasets). OPN and SHS parameter estimates are plotted on the X and Y axes, respectively. 90 (bold) and 95% (not bold) confidence intervals are shown.