

Fullard, Joshua; Sen, Sonkurt

Working Paper

Tell me who you are and I will give you my consent: A light-touch intervention on consent to data linkage

ISER Working Paper Series, No. 2022-10

Provided in Cooperation with:

Institute for Social and Economic Research (ISER), University of Essex

Suggested Citation: Fullard, Joshua; Sen, Sonkurt (2022) : Tell me who you are and I will give you my consent: A light-touch intervention on consent to data linkage, ISER Working Paper Series, No. 2022-10, University of Essex, Institute for Social and Economic Research (ISER), Colchester

This Version is available at:

<https://hdl.handle.net/10419/284112>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Tell Me Who You Are and I Will Give You My Consent: A Light-touch Intervention on Consent to Data Linkage

Joshua Fullard
Warwick Business School
University of Warwick

Sonkurt Sen
Department of Economics
University of Bonn

No. 2022-10
October 2022



INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

Non-technical Summary

Linking administrative data to survey data can be a powerful tool for researchers to maximise the value of their data without increasing respondent burden. However, it is often legally necessary to obtain a survey respondents' informed consent before linking their survey data to administrative records. Since not all respondents give their consent, this may be a source of bias if those who consent are systematically different than those who do not. This paper contributes to a literature that investigates how researchers can design surveys to maximise consent rates

Some of our earlier work suggests that participants trust in the research team plays an important role in participants decision to consent to data linkage (or not). In this paper we test this relationship directly by investigating: first, how participants trust in the research team differ by observable characteristics and second, the relationship between respondents trust in the research team and decision to consent to data linkage. Finally, we use a light touch randomised control trial to investigate how the provision of different sources of information (visual and/or written) about the research team involved in the study influences participants trust in the research team and their decision to consent to data linkage.

Our main results are as follows: first we find significantly lower levels of trust in the research team among black participants. Second, we find a strong positive relationship between respondents trust in the research team and their decision to consent to data linkage. Finally, looking at our information intervention we find that the provision of additional information about the research team has a positive effect on respondents' trust, particularly those from an ethnic minority background. However, the additional information about the research team has no effect on respondents consent rates. Respondents who increase their trust in the research team in response to the light touch information intervention are no more likely to consent to data linkage than those who did not.

Tell Me Who You Are and I Will Give You My Consent: A Light-touch Intervention on Consent to Data Linkage*

Joshua Fullard^{1,2} and Sonkurt Sen³

¹University of Warwick

²University of Essex

³University of Bonn

September 30, 2022

Abstract

Consent to data linkage (or the lack of it) has long been an important topic for researchers using survey data. In this paper, we use newly collected survey data to investigate the role that respondents' trust in the research team has on their propensity to consent to link their survey data to their educational, health and recreational administrative records. In addition, we investigate how participant consent rates differ by linkage topic and respondent characteristics. Finally, we experimentally vary the information respondents receive about the research team to investigate the role that information has on trust levels and consent rates. Our results show a strong positive relationship between respondents' trust in the research team and their propensity to consent to data linkage. Additionally, we find that the provision of additional information about the research team has a positive effect on respondents' trust, particularly those from an ethnic minority background. However, this increase in the respondents trust does not translate into higher consent rates.

Keywords: Data Linkage, Consent, Light-touch Intervention

*This project was funded by ESSEXLab through Seedcorn Grant scheme. We are grateful for the comments received from Peter Lynn and other colleagues at the University of Essex. Corresponding author: Fullard, Warwick Business School, University of Warwick, Joshua.Fullard@wbs.ac.uk

1 Introduction

Due to technological advances and the development of new empirical techniques, researchers have many potential new data sources that can enhance survey data (e.g. social media data, education records and pension records). While these data sources are becoming increasingly available to researchers, linking this data to survey data often requires respondents' informed consent to the linkage. In many countries obtaining informed consent is even a legal requirement - in the UK, this is covered by the Digital Economy Act. While obtaining informed consent might not always be required from a legal perspective, it is often important for researchers from an ethical and practical perspective - many ethical review boards will require respondents' informed consent. Since not all respondents give their consent, this may create a source of bias if the respondents who consent are systematically different than those who do not. While there are empirical tools to correct for this bias, there is a growing literature in Survey Methodology that studies how consent rates differ by participant characteristics and how researchers can design surveys to maximize consent rates.

The findings of the literature generally shows that a range of factors can influence the decision for a participant to consent to data linkage, or not. These factors include the mode of interview: [Jäckle *et al.* \(2021b\)](#) find that web respondents have significantly lower consent rates (42 percent) compared to face-to-face respondents (73 percent) which is supported by [Sakshaug *et al.* \(2017\)](#) and [Thornby *et al.* \(2018\)](#) who find online consent rates of 54 percent (vs 95 percent face-to-face) and 69 percent (vs 89 percent for face-to-face) respectively. The type of data that is to be linked is also known to influence consent to data linkage. Recently, [Edwards & Biddle \(2021\)](#) find that consent rates to educational and health records were significantly higher than to income or employment records. This is similar to [Walzenbach *et al.* \(2022\)](#) who find higher rates of consent to health records (52 percent) than tax records (48 percent).

The existing literature also shows that question framing influences the decision to consent, particularly in the web based surveys. [Kreuter *et al.* \(2016\)](#) and [Sakshaug & Kreuter \(2014\)](#) find that positive framing - emphasizing the benefits of data linkage to respondents - has a positive effect on obtaining consent. However, there is limited evidence that framing has any effect on

consent to data-linkage in interviewer-administered surveys (Pascale, 2011; Sakshaug *et al.*, 2013). Other factors that are thought to influence consent rates include the timing of the question. Most surveys generally have the request to data linkage at the end of the survey, but empirical evidence suggests that this lowers consent rates. Using telephone interviews in Germany, Sakshaug *et al.* (2013) find that placing the request at the start of the survey achieves higher consent rates than at the end (96 percent vs 86 percent). This is supported by Sala *et al.* (2014) who find that, in a UK based web survey, respondents are more likely to consent to data linkage when the request is placed earlier in the survey (65 percent) than at the end (58 percent). But the literature is not conclusive as Jäckle *et al.* (2021b) suggest that as long as the question is easy to understand the placement has no impact on consent rates.

Beyond the survey literature, a growing medical literature has found that individuals are often not fully informed about medical decisions but rarely use the available resources to improve their understanding (Douglas *et al.*, 2021; McNutt *et al.*, 2008). Despite the fact that individuals often have very poor understanding of the medical decisions they are faced with, subjects generally feel informed and confident about the decisions they are making - possibly suggesting why the provision of additional information rarely increases consent rates in the survey literature (Fagerlin *et al.*, 2010; Hoffmann & Del Mar, 2015). While individuals' beliefs about their own understanding is often unrelated to their actual understanding, Sepucha *et al.* (2010) finds that they are strongly related to the trust they have in their doctor - if respondents trust their doctor, they are more likely to report that they are well informed and have higher confidence that they have made the best possible decision.

Building on this medical literature, empirical evidence shows that participant trust is an important driver of consent to data linkage. Individuals who are more trustworthy, in general, are significantly more likely to consent to data linkage (Sala *et al.*, 2012). Furthermore, the empirical evidence shows that trust at the institutional level is strongly correlated with the respondents' decision to consent to the data linkage. More recently, Jäckle *et al.* (2021b) finds that respondents who said that they trusted HM Revenue and Customs (HMRC) were more likely to consent to have

their data linked to HMRC records than those who did not (58 percent compared to 35 percent). Recent evidence also shows that participants' trust in the research team influences consent to data linkage. Using a survey of teachers, [Fullard \(2022\)](#) finds that teachers from schools that have a connection to the university that the research team is based in have a higher propensity to consent to link their survey data with their employment records (81 percent vs 73 percent). The author argues that the higher propensity to consent to data linkage is driven by a greater trust in the research team. In this paper, we are able to test this hypothesis directly.

In this paper, we built on the current literature and address the following four main research questions: First, how do respondents' trust in the research team differ by observable characteristics. Second, how do consent rates differ by topic, participants observable characteristics, and how do these rates compare to existing empirical studies. Third, how do respondents' trust in the research team relate to their decision to consent to data linkage. We then extend this research question and investigate whether the topic of the linkage is important, i.e. we study consent decisions, and the role of trust, in three domains: health, education and tax records, and differences by respondent characteristics. Fourth, using a randomized controlled trial, we investigate how the provision of information about the primary researcher involved in this study influences participants' trust in the research team and their propensity to consent to data linkage. Here, our treatment contains three arms: In the first arm, we show the participants only the photo of the primary researcher smiling (visual information only), in the second treatment arm, we provide the participants with information about the academic qualifications of the researcher as well as the publicity of their research (written information only). In the third treatment arm, we combine the information about the researcher with their photo (visual and written information) while our control group received neither the information nor the photo of the primary researcher.

Recall that the existing literature generally finds higher consent rates in face-to-face surveys, compared to web surveys. One reason for this is the rapport between an interviewer and respondent in a face-to-face survey that is not present in a web survey. By experimentally varying the information participants have about the research team - a picture of the primary researcher smil-

ing and/or the primary researcher’s achievements and experience - we are able to investigate the effect of visual and written information, and how these two effects might combine, in the context of participant trust levels and subsequent consent rates in web surveys. The fact that this is a light touch intervention, providing information in different formats (visual vs written), is an important contribution of this study as many of the interventions that was implemented previously to improve consent rates are expensive (such as providing financial incentives, interview follow-ups, extra mailings, etc) so investigating the impact of a cheap, easy to implement, information intervention could be a significant advancement to the survey qualities and what researchers can achieve.

Our four main results are as follows: First, we observe significantly lower rates of trust in the research team among Black participants. Second, consent rates in our study are generally lower than other web-based studies - possibly reflecting the sensitive nature of some of the topics we collect data on (e.g., racial biases and performance evaluations). Consistent with some of the existing literature, we find that consent rates are highest for data linkage to educational records and lowest for health records. We also find a significant differences in consent rates by ethnicity which highlights that there are likely to be systematic differences among those who do, and do not, consent to data linkage which is important for researchers using survey data as the external validity of their results can be harmed by this. Third, we find a strong positive relationship between respondents’ trust in the research team and their propensity to consent to data linkage. A higher trust in the research team is associated with a higher rate of consent to data linkage. Fourth, looking at our information intervention, we find that the provision of additional information about the research team has a positive effect on respondents’ trust, particularly those from an ethnic minority backgrounds. However, the additional information about the research team has no effect on respondents’ consent rates. This is not necessarily surprising given that the change in respondents trust in the research team is uncorrelated with their propensity to consent to data linkage. Respondents who increase their trust in the research team in response to the light touch information intervention are no more likely to consent to data linkage than those who did not.

One of the unique contributions of this paper is that we investigate the effect that the provision of a light touch information intervention about the research team has on respondents' trust in the research team and respondents propensity to consent to data linkage. This contributes to a body of work seeking to determine whether the provision of information improves consent rates. While the provision of additional information is commonly used to increase consent rates, the effectiveness of this approach is not well established. There is some evidence that the provision of additional information does improve consent rates such as [Graves *et al.* \(2019\)](#), using a web administered survey of women in Australia, and [Fullard \(2022\)](#) in a web administered survey of teachers in the UK. The literature predominately finds that additional information does not increase consent rates ([Jäckle *et al.*, 2021a](#)) possibly because the additional information provided is too generic and therefore does not address the specific concerns a participant might have in the same way an interviewer could in an interviewer-administered survey. Alternatively, participants might believe that they are already fully informed and don't feel the need to access additional information.

Our paper is organized as follows: Section 2 provides the study design, Section 3 explains the empirical strategy, Section 4 present the results and section 5 concludes.

2 Study Design

The data we use comes from a larger study studying how people evaluate performance. This study has 6 sections: i) Demographic survey, ii) Implicit Association Test and attitudes towards racial minorities, iii) Attractiveness rankings, iv) Performance evaluation, v) Team building exercise and vi) Trust in research team and consent to data linkage. In this paper, we use data from the first and last section to study our main research questions. In the following subsection we introduce the overall design (sample, data collection mode, timings etc.) followed by a discussion of each of our measurement instruments.

2.1 Sample and Randomization

The study was designed as a web-survey and the treatment was administrated online. The invitations for the study was sent four times in June 2022 and the data collection was finalized by the end of that month. Respondents were recruited using the university’s social science laboratory’s online subject pool – individuals need to have signed up to take part in social science experiments in order to receive invitation for our study. One section of the survey had to be completed on a computer so we restricted the sample to those who can complete the survey on a computer (and not a mobile device) and, as another part of the study required the respondents to watch some video materials that can only be accessed in the UK, the sample is restricted to those who were located in the UK at the time of their completion. Overall, the survey took around 20 minutes to complete, including the other parts of the study, and the respondents were paid a small amount as a thank you for taking part (between 3 GBP and 7 GBP. The average payment was 5 GBP), and the amount was calculated based on the responses given in other sections of the questionnaire, not relevant to this study.

We present the characteristics of the respondents in [Table 1](#). Our main sample consists of 519 individuals. The majority of our respondents are female (63%) and the average age of our sample is 28.63 years. While this might seem like a very old age as the invitations were sent from the social science lab of a university, the lab allows non-students to register to their participant pool meaning anyone with an email address can register. Indeed, we see that 36% of our sample consists of non-students who might be staff members at the university or members of public. When we look at the racial composition of our sample, we see that ethnic minority participants are over-represented compared to the population of the UK. While only 3.3 and 7.5% of the UK population are Black and Asians, 12 and 20% of our sample are individuals from Black and Asian racial groups. This is, however, not surprising given the university that this study conducted advertises itself as an international university and its student population consists of students from many countries. Similarly, our sample’s maternal education levels are high compared to the UK population but as university enrollment is correlated with parental education, this is expected.

Randomization to treatment and control groups was performed by the survey software. The information intervention consists of 3 treatment arms and 1 control group. In columns 2-8 of [Table 1](#), we show the characteristics of the individuals in each of these conditions and the p-value for the differences between each treatment arms and the control group. While there are some differences in personal characteristics between each treatment arm and the control group, we find that none of these differences are statistically different except maternal education between control group and Photo treatment. While Photo treatment group have more people whose mothers have more likely to have A-levels (high school degree) but less likely to have a postgraduate degree, these differences are only weakly significant. While this is unlikely to bias the results in a meaningful way, there is also a chance that it might still affect the results. In order to avoid this possible bias, we control for maternal education in our main empirical specification. In [Table 2](#), we show the balance of baseline trust between each treatment arms and control group and the differences by gender and ethnicity as well as differences by end-line consent to data linkage variables. [Table 2](#) complements [Table 1](#) in showing that our randomization was successful in randomizing the sample so that the baseline trust in the research team across the treatment groups and the control group is balanced. These two tables ensure that when we study the effect of our intervention on the trust in research team and the consent to data linkage variables, our results will show the unbiased and causal treatment effect.

2.2 Trust in the Research Team

We measure respondents trust in the research team twice, before (in section iii of the survey) and after our intervention (in section vi). Although there are several questionnaires to measure the trust of individuals such as the one proposed by [Yamagishi & Yamagishi \(1994\)](#), we follow a different approach. As our trust in this paper concerns with the trust in the research team, we follow a method that is more commonly implemented in the surveys, i.e. asking the respondents directly how much they trust the research team following [Fehr *et al.* \(2003\)](#) and [Naef & Schupp \(2009\)](#).

An alternative method of measuring participants trust, or any other belief, is by providing participants with several statements (e.g. the research team is trustworthy) and asking participants how much they agree with each statement. As there is empirical evidence from the Survey Methodology literature that clear and direct questions improve data quality in certain settings, such as consent to data linkage (Jäckle *et al.*, 2021b), it is important to avoid any possibility of confusion which is why we ask participants about their trust directly in a clear and easy to understand way. Although not directly related, there is also evidence that asking questions directly, rather than experimentally eliciting beliefs or preferences, is one of the appropriate ways to avoid confusion and improve the quality of the data (Dohmen *et al.*, 2011).

In order to measure trust in the research team respondents were asked the following questions where respondents provided their answers on a 0-100 scale using a sliding bar.

On a scale from 0% to 100%, could you please tell me how much you trust the research team. 0% means you do not trust the research team at all and 100% means you trust the research team completely.

We use the baseline measure (pre-intervention from section III) to check for balance across treatment groups and to study the relationship between the trust of respondents and their likelihood of giving consent. Then, we use both the baseline measure and end-line measure (post-intervention measure from section VI) to study the effect of our intervention on respondents' trust level.

2.3 Consent to Data Linkage

After the trust question in section VI of the survey, we also ask the participants whether they would be willing to give consent to data linkage. We asked this question in three domains: education records (link to their university attendance records and their grades), health records (held by National Health Service) and their university activities (sports club, student clubs membership etc.). In order to avoid any potential ordering effects, the order of the topics that we asked for consent was randomized by the survey software.

While we ask trust in research team questions both in the baseline and end-line, we ask consent

to data linkage questions only in the end-line. This is due to time constraints of the main survey as it was a larger study and because asking all three consent questions twice would raise the possibility of experimenter demand effect, i.e. respondents guessing what the treatment is and answering accordingly or respondents feeling pressured to giving their consent which might also bias the results. Relatedly, since by law it is only required to receive consent from the participants once, asking these questions twice would raise the possibility of the suspicion of the respondents, thus rendering the data quality of the survey. The exact wording of the questions are as follows:

University Records – *We would like to link the answers you have given in this study to your attendance records and marks held by the university. Do you give permission for us to pass your details to the university for this purpose?*

Health Records – *We would like to link your answers to your health records held by the National Health Service (NHS). Do you give permission for us to pass your details to the NHS for this purpose?*

Student Activities – *We would like to link your answers to your sports clubs/societies membership(s) held by the Students Union. Do you give permission for us to pass your details to the SU for this purpose?*

Before each of these questions, participants were warned that their decision to consent will not affect their participation in the survey nor they have any obligation to consent to data linkage.

Once the respondents completed the section VI trust and consent to data linkage questions, they were told that no data linkage will actually take place. As some of our consent questions are context-dependent such as academic outcomes and university activities, they will be more relevant to some participants than others, i.e. the current students at the study university. For this reason, in our results, we also study the effects of the treatment on consent to data linkage both for our entire sample and for the sub-sample that consists of only students.

2.4 Intervention

At the start of section VI, participants were randomly allocated to one of the intervention groups. Our study has 4 groups in total, 1 control and 3 treatment groups. Randomization is needed to study the casual effect of the information on the participants' trust in the research team and the consent to linkage. It is especially important for the consent to linkage part because, as mentioned in the previous section, we only collect participants' consent once. If we did not randomized the intervention, we would not be able to study the causal effect of our intervention on consent rates as there might be other differences between treatment and control groups (both observable and unobservable differences) that would affect our outcome of interest. However, by randomizing our intervention, we ensure that the only difference between the groups is the provision of information, therefore we can cleanly identify the casual effect of the information experiment on the participants' trust in the research team as well as on their likelihood of providing consent for data linkage.

Our study consists of three different but related treatment arms. In the first arm, Photograph Treatment, the respondents were shown the photo of the principal researcher smiling in a suit to show that the researcher is a friendly person. In the second arm, Information Treatment, we provided the respondents with the academic qualification of the principal researcher as well as some information about the publicity of their research to show that the researcher is a trustworthy individual. In the third arm, we combined the first arms and showed both the photo of the principal researcher smiling and information about their academic qualifications as well as the publicity of their research. Participant who were allocated into the control group received neither the photo of the principal researcher nor information about their academic qualifications and their research (except when they arrived to the end of the survey, they were told who to contact in case of a question but they had already made their decisions about their trust and consent at that point). The information that the different treatment groups were provided is available in the Appendix B.

3 Empirical Strategy

We estimate the effect of our intervention first on the trust that the respondents have about the research team. As mentioned in the previous sections, we have information about individuals' trust in the research team before they received the intervention. We make use of this information to improve the precision of our estimates of the effect of the intervention on the trust in the research team and estimate the following specification:

$$y_{it} = \alpha + \beta T_i + \theta y_{it-1} + \gamma \mathbf{X}_i + \epsilon_{it} \quad (1)$$

where y_{it-1} and y_{it} are trust in research team before and after the intervention, T_i is the intervention group and the type of treatment the individuals received if in the treatment group, \mathbf{X}_i is the vector of respondent characteristics which includes gender, age, whether the respondent is student, maternal education and ethnicity. Here, β gives us the causal effect of the intervention on the trust in research team.

We, then, move on to estimating the effect of our intervention on the consent to data linkage for the three domains that we study. However, as we do not have a measure of consent-giving before our intervention, we cannot estimate the above model. While one might think that this is a problem, many of the RCTs in the literature do not have a baseline measure for their outcomes of interest. As RCTs ensure that the observable and unobservable characteristics of the respondents allocated to each group would be balanced, the literature generally studies the effect of RCTs without controlling for the baseline measure of the outcome of the interest. We follow the literature and estimate the following model to study the effect of our intervention on the consent to data linkage:

$$y_{it} = \alpha + \beta T_i + \omega Trust_i + \gamma \mathbf{X}_i + \epsilon_{it} \quad (2)$$

where y_{it} is whether the participant is willing to give consent to data linkage and $Trust$ is baseline trust. Here, we estimate this equation three times for three outcomes: consent to data linkage to

i) academic records, ii) health records and iii) student activity records.

4 Results

4.1 Trust in the Research Team

We first describe how participants' trust in the research team correlates with personal characteristics in [Table 3](#). Here, we use our baseline measure of trust as this is the measure that is not changed by our intervention. We observe that while gender, being a student and age are not correlated with the trust levels, there are some differences by maternal education and ethnicity. First, we find that Black respondents, on average, report lower trust levels than White respondents. While the mean trust level is 74.6, Black individuals are reporting 7.04 points lower levels of trust in the research team. While this might seem like a small difference compared to the mean, it is 32% of the standard deviation, showing that this is a meaningful difference. While this is consistent with the literature that ethnic minorities tend to hold lower levels of societal trust ([Ziller, 2017](#)), it is surprising that there are no other statistically significant differences between our White respondents and those from the other ethnic minority groups. Second, we find that those who have a mother with a university degree have higher levels of trust than those whose mother has only completed education up until 16. Similar to the ethnic difference, the difference is 6.01 points and this is 28% of the standard deviation, showing that the difference of having a mother with university degree is quite strong. We also observe striking differences among those who do not have a mother, or female caregiver. These individuals are 20 percentage points less likely to trust the research team than those with the least educated mothers –although this is a very small group (15 respondents do not have a mother or female caregiver) so this result should be interpreted with caution.

4.2 Consent to Data Linkage

We next describe who gives consent to data linkage. We focus on three contexts: i) Education Records, ii) Health Records and iii) Student Activity Records. In [Table 2](#), we show the mean

consent rate for all three context first for the whole sample, and then by the treatment arm and finally by gender and race. We find significant variation in the consent rates. We find that 54% of the individuals give consent to linkage to education records while 46 and 50% of respondents consent to linkage to health and student activity records, respectively. While the ordering of these topics is consistent with other empirical findings - [Edwards & Biddle \(2021\)](#) find that consent rates to educational records were higher than health records which is what we also find – the rates of consent in our sample are generally lower than the existing literature. For example, [Walzenbach et al. \(2022\)](#) and [Graves et al. \(2019\)](#) find consent rates of 52 percent and 69 percent respectively to link survey data to health records (compared to our consent rate of 46 percent). [Al Baghal \(2016\)](#) and [Thornby et al. \(2018\)](#) find consent rates to education records of 70 percent and 63 percent respectively (compared to our consent rate of 54 percent). These differences are likely to be driven by some combination of compositional differences across the samples and the fact that our survey collects data on sensitive topics (e.g., measures of racial biases) that might make respondents less inclined to consent to linkage.¹

Looking at the relationship between consent to data linkage and our respondents’ background characteristics in [Table 5](#), three main things stand out. First, Black respondents are significantly more likely to consent to data linkage than their White counterparts. Moreover, the magnitude is large. Black respondents are 21 percentage points more likely to consent to data linkage with their health records than White respondents (column 2). This is surprising as the existing literature has generally found that ethnic minority groups are less likely to consent to data linkage ([Mostafa, 2016](#); [Woolf et al., 2000](#)). Moreover, Black respondents have slightly lower levels of trust in the research team which are thought to be an important driver in consent to data linkage.

Second, we find no differences in consent rates by age for linkage to education records, but we do observe a negative effect for linkage to student activity and health records, although the magnitude is small - an increase in respondents age from the median (24) to the 75th percentile

¹While this is not a direct comparison to our consent rates [Sakshaug et al. \(2017\)](#) find a consent rate of 54 percent and [Thornby et al. \(2018\)](#) find a consent rate of 69 percent to employment records. Other factors such as placement might also explain the lower consent rate in our setting.

(31) is associated with a 4.2 percentage point lower consent rate. This is similar to the existing literature which generally finds modest differences by age (Bryant *et al.*, 2006). Third, turning our attention to sex we also observe no differences in consent rates. While this is consistent with some studies (Klassen *et al.*, 2005) the literature generally finds higher rates of consent for men (Huang *et al.*, 2007).

Finally, students are more likely to consent to data linkage with respect to their education and student activity records than non-students although this might be driven by the fact that the linkage to educational and student activity records might be more relevant for current students. Indeed, there is no difference in the propensity to consent for health records among students and non-students.

There is also evidence that trust in the research team is an important driver in the decision to consent to data linkage. We investigate this directly by looking at the relationship between the respondents' trust and propensity to consent to data linkage. We observe that respondents' baseline trust in the research team is positively associated with their propensity to consent to data linkage. A 3.5 percentage point increase in trust (the average treatment effect we observed in the previous section) is associated to a 1.40 percentage point higher likelihood that a participant will consent to linking their survey data with their educational records (column 1), a 1.75 percentage point higher likelihood for linkage to recreation records (column 2) and for linkage to health records (column 3). These differences are not statistically significant.

Interestingly, we find that the respondents' revised trust in the research team is not related to their propensity to consent to data linkage - further analysis shows that the change in the respondents' trust in the research team has, a precisely estimated, no effect on their propensity to consent to data linkage (not reported but available on request). Respondents who increase their trust in the research team are no more likely to consent to data linkage than those who did not.

4.3 Information Treatment

Turning our attention to the effect of the information intervention, we first study the effect on participants trust in the research team. In [Table 4](#), we show the results for the whole sample in columns 1-3 and then for sub-samples. In column 1, we only have intervention dummies while in columns 2 and 3, we include the baseline and then personal characteristics. As we can see from the table, the inclusion of the personal characteristics does not make much of a difference in the magnitude or the significance of the effects but it is still important to include as we have observed some differences in the maternal education across different treatment groups. Our main hypothesis related to trust is that providing information about and/or photograph of the principal research would improve the trust levels because it shows that the researcher has a strong track record of research and that they are a friendly person, in line with Ethos of Aristotle’s Trinity of Persuasion. Indeed, this is what we observe in column 3 of [Table 4](#). We see that providing information about the researcher or the photo of the researcher does not have any effect on the trust level of the individuals. However, when we combine these two treatments, we see that individuals increase their trust level by 3.55 points (or 16% of standard deviation) compared to the control group. The finding that the picture treatment had no impact on respondents’ level of trust is unsurprising as a picture without any context (e.g., details about the research team) is unlikely to communicate anything tangible to the respondent – if anything it might come across as confusing or even suspicious.

Despite our relatively small sample size, columns 4–9 of [Table 4](#) investigates the heterogeneity in treatment effect by sub-sample: i) student status, ii) gender and iii) race. We find that the treatment effect is entirely coming from students, female individuals and minorities. We also find that the effect of the Information Treatment is heterogeneous by ethnicity (significant at 5% significance level), ie. those from ethnic minority backgrounds are more affected from this treatment. This is perhaps not surprising given that they had lower levels of trust to begin with. Although the differences between each of the other groups are not statistically significant (possibly due to low sample sizes), we find these effects interesting. While the results on minority

individuals are consistent with the initial level of trust, the heterogeneity by gender and student status is interesting. Our results suggest that the treatment effect comes entirely from females and students. However, [Table 3](#) shows that there is no gender differences in initial trust levels nor there is when it comes to being a student or not so we can rule out the possibility of more "room to improve" when it comes to these sub-samples.

We, next, explore how the light touch information treatments impacts respondents' propensity to consent to data linkage. [Table 6](#) reports the coefficients associated with the treatment groups (compared to the control group) for consent to linkage to education records (columns 1-3), student activity records (columns 4-6) and health records (columns 7-9). We find that our intervention does not have any effect on the consent to linkage of records in all of our contexts.

Recall that i) data linkage might be more relevant to some respondents than others - the linkage of data to education and student activity records might be more relevant to current students - and ii) larger treatment effects for non-white respondents trust in the research team. In [Table 6](#) we also investigate the heterogeneity in treatment effect by restricting our sample to current students and minority participants . Similar to our main results, we find no treatment effect for education or health linkage and in fact some of the coefficients are very close to 0 although none of the other coefficients are statistically different than 0. However, we observe an effect for student activity activity records, and the magnitude is large. Minority participants who received the picture only treatment are 18.1 percentage points (or 36% of the standard deviation) less likely to consent to data linkage to student activity records than the control group. While we are not necessarily surprised that the Photo treatment has a negative effect – seeing someone's pictures and then asked to consent to data linkage could be seen as suspicious – we are surprised by both the magnitude and the fact that it is not consistent across topics, it is not clear why this treatment would only affect consent to student activity records and not educational and health records. Therefore, this result should be treated with caution.

5 Conclusion

Linking administrative data to survey data can be a powerful tool for researchers. However, it is often legally necessary to obtain a survey respondents' informed consent before linking their survey data to administrative records. Since not all respondents give their consent, this may be a source of bias if those who consent are systematically different than those who do not. This paper studies the relationship between respondents' trust in the research team and their propensity to consent to data linkage and how consent rates differ by topic and participant characteristics. Finally, we investigate the effect of a light touch information intervention, about the research team, has on respondents' trust in the research team and respondents' propensity to consent to data linkage.

Our four main results are as follows: first, we observe significantly lower rates of trust in the research team among black participants. Second, consent rates in our study are generally lower than other web-based studies - possibly reflecting the sensitive nature of some of the topics we collect data on (e.g., racial biases and performance evaluations). Consistent with some of the existing literature, we find that consent rates are highest for data linkage to educational records and lowest for health records. We also find a significant difference in consent rates by ethnicity which highlights that there are likely to be systematic differences among those who do, and do not, consent to data linkage which is important for researchers using survey data as the external validity of their results can be harmed by this. Third, we find a strong positive relationship between respondents' trust in the research team and their propensity to consent to data linkage. A higher trust in the research team is associated to a higher rate of consent to data linkage. Fourth, looking at our information intervention we find that the provision of additional information about the research team has a positive effect on respondents' trust, particularly those from an ethnic minority background. However, the additional information about the research team has no effect on respondents' consent rates. This is not necessarily surprising given that the change in respondents trust in the research team is uncorrelated with their propensity to consent to data linkage. Respondents who increase their trust in the research team in response to the light touch information intervention are no more likely to consent to data linkage than those who did not.

Our interpretation of these results is that while the provision of information about the research team may increase respondents' trust in the research team it is still a very light touch intervention. Influencing the respondents' decision to consent to data linkage, or not, might require a more specific intervention (e.g., information about the research teams experience in handling data linkage), or something more powerful (e.g., a video of the research team discussing the values of data linkage). In a wider context, these result provide further evidence that the provision of information can be an inexpensive approach to influence participants' beliefs and expectations. However, the type of additional information does need to be carefully considered - heterogeneity analysis suggests that providing participants with a smiling picture of the primary researcher actually reduced consent rates - the picture without context might have come across as suspicious.

This paper takes an important step in investigating how respondents' trust in the research team, and the provision of information about the research team, affects respondents' propensity to consent to data linkage. However, more work needs to be done. Investigating the impact of a stronger and/or more specific information intervention, such as a video of the research team discussing the values of data linkage, seems like a promising areas of future research. In addition, exploring how reference points influences consent rates also seems promising such allocating respondents to different reference points (e.g., in [Fullard \(2022\)](#) where 75 percent of respondents consented or [Sakshaug *et al.* \(2017\)](#) where 54 percent consented to data linkage).

References

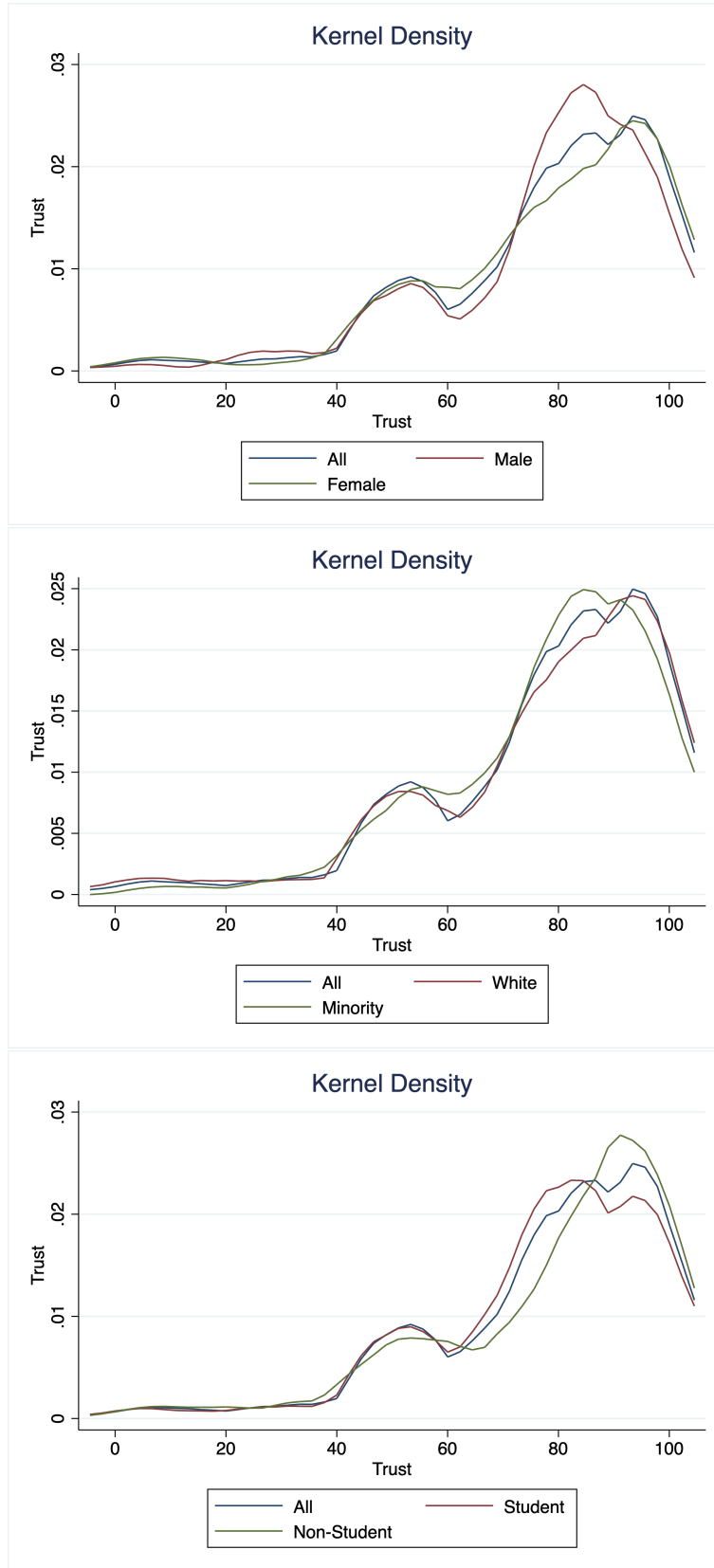
- Al Baghal, Tarek. 2016. Obtaining data linkage consent for children: factors influencing outcomes and potential biases. *International Journal of Social Research Methodology*, **19**(6), 623–643.
- Bruhn, Miriam, & McKenzie, David. 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, **1**(4), 200–232.
- Bryant, Heather, Robson, Paula J, Ullman, Ruth, Friedenreich, Christine, & Dawe, Ursula. 2006. Population-based cohort development in Alberta, Canada: a feasibility study. *Chronic Dis Can*, **27**(2), 51–59.
- Dohmen, Thomas, Falk, Armin, Huffman, David, Sunde, Uwe, Schupp, Jürgen, & Wagner, Gert G. 2011. Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences. *Journal of the european economic association*, **9**(3), 522–550.
- Douglas, Benjamin D, McGorray, Emma L, & Ewell, Patrick J. 2021. Some researchers wear yellow pants, but even fewer participants read consent forms: Exploring and improving consent form reading in human subjects research. *Psychological Methods*, **26**(1), 61.
- Edwards, Ben, & Biddle, Nicholas. 2021. Consent to Data Linkage: Experimental Evidence from an Online Panel. *Advances in Longitudinal Survey Methodology*, 181–203.
- Fagerlin, Angela, Sepucha, Karen R, Couper, Mick P, Levin, Carrie A, Singer, Eleanor, & Zikmund-Fisher, Brian J. 2010. Patients' knowledge about 9 common health conditions: the DECISIONS survey. *Medical Decision Making*, **30**(5_suppl), 35–52.
- Fehr, Ernst, Fischbacher, Urs, Von Rosenbladt, Bernhard, Schupp, Jürgen, & Wagner, Gert G. 2003. A nation-wide laboratory: Examining trust and trustworthiness by integrating behavioral experiments into representative survey. *DIW Discussion Paper*.

- Fullard, Joshua. 2022. *An information intervention and consent to data linkage: experimental evidence from teaching*. Tech. rept. Institute for Social and Economic Research.
- Graves, Anna, McLaughlin, Deirdre, Leung, Janni, & Powers, Jennifer. 2019. Consent to data linkage in a large online epidemiological survey of 18–23 year old Australian women in 2012–13. *BMC medical research methodology*, **19**(1), 1–9.
- Hoffmann, Tammy C, & Del Mar, Chris. 2015. Patients’ expectations of the benefits and harms of treatments, screening, and tests: a systematic review. *JAMA internal medicine*, **175**(2), 274–286.
- Huang, Nicole, Shih, Shu-Fang, Chang, Hsing-Yi, & Chou, Yiing-Jenq. 2007. Record linkage research and informed consent: who consents? *BMC Health Services Research*, **7**(1), 1–5.
- Jäckle, Annette, Burton, JONATHAN, Couper, Mick P, Crossley, Thomas F, & Walzenbach, SANDRA. 2021a. *How and why does the mode of data collection affect consent to data linkage?* Tech. rept. Understanding Society Working Paper 2021-04, Colchester: University of Essex
- Jäckle, Annette, Burton, Jonathan, Couper, Mick P, Crossley, Thomas F, & Walzenbach, Sandra. 2021b. Understanding and improving data linkage consent in surveys.
- Klassen, Anne F, Lee, Shoo K, Barer, Morris, & Raina, Parminder. 2005. Linking survey data with administrative health information. *Canadian journal of public health*, **96**(2), 151–154.
- Kreuter, Frauke, Sakshaug, Joseph W, & Tourangeau, Roger. 2016. The framing of the record linkage consent question. *International Journal of Public Opinion Research*, **28**(1), 142–152.
- McKenzie, David. 2012. Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, **99**(2), 210–221.
- McNutt, Louise-Anne, Waltermaurer, Eve, Bednarczyk, Robert A, Carlson, Bonnie E, Kotval, Jerro, McCauley, Jeanne, Campbell, Jacquelyn C, & Ford, Daniel E. 2008. Are we misjudging

- how well informed consent forms are read? *Journal of Empirical Research on Human Research Ethics*, **3**(1), 89–97.
- Mostafa, Tarek. 2016. Variation within households in consent to link survey data to administrative records: evidence from the UK millennium cohort study. *International Journal of Social Research Methodology*, **19**(3), 355–375.
- Naef, Michael, & Schupp, Jürgen. 2009. Measuring Trust: Experiments and Surveys in Contrast and Combination. *IZA Working Paper*.
- Pascale, Joanne. 2011. Requesting consent to link survey data to administrative records: Results from a split-ballot experiment in the Survey of Health Insurance and Program Participation (SHIPP). *Survey Methodology*, **3**.
- Sakshaug, Joseph, Tutz, Valerie, & Kreuter, Frauke. 2013. Placement, wording, and interviewers: identifying correlates of consent to link survey and administrative data. *Pages 133–144 of: Survey Research Methods*, vol. 7.
- Sakshaug, Joseph W, & Kreuter, Frauke. 2014. The effect of benefit wording on consent to link survey and administrative records in a web survey. *Public Opinion Quarterly*, **78**(1), 166–176.
- Sakshaug, Joseph W, Hülle, Sebastian, Schmucker, Alexandra, & Liebig, Stefan. 2017. Exploring the effects of interviewer-and self-administered survey modes on record linkage consent rates and bias. *Pages 171–188 of: Survey Research Methods*, vol. 11.
- Sala, Emanuela, Burton, Jonathan, & Knies, Gundi. 2012. Correlates of obtaining informed consent to data linkage: respondent, interview, and interviewer characteristics. *Sociological Methods & Research*, **41**(3), 414–439.
- Sala, Emanuela, Knies, Gundi, & Burton, Jonathan. 2014. Propensity to consent to data linkage: experimental evidence on the role of three survey design features in a UK longitudinal panel. *International Journal of Social Research Methodology*, **17**(5), 455–473.

- Sepucha, Karen R, Fagerlin, Angela, Couper, Mick P, Levin, Carrie A, Singer, Eleanor, & Zikmund-Fisher, Brian J. 2010. How does feeling informed relate to being informed? The DECISIONS survey. *Medical Decision Making*, **30**(5-suppl), 77–84.
- Thornby, Marie, Calderwood, Lisa, Kotecha, Mehul, Beninger, Kelsey, & Gaia, Alessandra. 2018. Collecting Multiple Data Linkage Consents in a Mixed-mode Survey: Evidence from a large-scale longitudinal study in the UK. *Survey methods: Insights from the field*, 1–14.
- Walzenbach, Sandra, Burton, Jonathan, Couper, Mick P, CROSSLEY, Thomas Fraser, & JÄckle, Annette. 2022. Experiments on multiple requests for consent to data linkage in surveys. *Journal of survey statistics and methodology*.
- Wolf, Steven H, Rothemich, Stephen F, Johnson, Robert E, & Marsland, David W. 2000. Selection bias from requiring patients to give consent to examine data for health services research. *Archives of family medicine*, **9**(10), 1111.
- Yamagishi, Toshio, & Yamagishi, Midori. 1994. Trust and commitment in the United States and Japan. *Motivation and emotion*, **18**(2), 129–166.
- Ziller, Conrad. 2017. Equal treatment regulations and ethnic minority social trust. *European Sociological Review*, **33**(4), 563–575.

Figure 1: Baseline Trust by Gender, Race and Student Status



Tables

Table 1: Descriptive Statistics and Balancing on Characteristics

Characteristics	All	Control	Information Treatment	p-value	Photograph Treatment	p-value	Combined Treatment	p-value
Male	0.37	0.37	0.30	0.29	0.42	0.53	0.44	0.35
Age	28.63	27.34	28.17	0.54	28.60	0.33	29.43	0.15
Student	0.64	0.62	0.62	0.99	0.67	0.39	0.66	0.47
Ethnicity								
White	0.58	0.58	0.63	0.46	0.55	0.69	0.55	0.66
Black	0.12	0.11	0.10	0.71	0.12	0.93	0.17	0.23
Asian	0.20	0.22	0.17	0.29	0.23	0.89	0.18	0.40
Other	0.10	0.09	0.11	0.60	0.11	0.71	0.11	0.66
Mother's Education								
GCSE	0.26	0.25	0.25	1.00	0.26	0.95	0.26	0.89
A-Levels	0.26	0.22	0.27	0.43	0.33*	0.09	0.23	0.93
Undergraduate	0.30	0.33	0.30	0.56	0.28	0.43	0.33	0.93
Postgraduate	0.14	0.17	0.15	0.64	0.09*	0.07	0.15	0.68
Missing	0.04	0.02	0.03	0.52	0.04	0.33	0.03	0.55
N	519	99	115		113		119	

Notes: p-values are for the differences between control condition and each of the treatment conditions. * denotes significance at 10% level, ** denotes significance at 5% level and *** denotes significance at 1% level where base condition is control.

Table 2: Descriptive Statistics and Balancing on Trust and Consent Rates

	All	Control	Information Treatment	Photograph Treatment	Combined Treatment	Male	Female	White	Minority
Baseline Trust	74.60 (21.79)	75.23 (21.84)	76.56 (21.09)	73.27 (21.62)	75.08 (22.84)	73.45 (22.21)	75.27 (21.55)	75.23 (21.78)	73.72 (21.83)
Education Linkage	0.54 (0.50)	0.55 (0.50)	0.58 (0.50)	0.49 (0.50)	0.53 (0.50)	0.54 (0.50)	0.54 (0.50)	0.52 (0.50)	0.56 (0.50)
Health Linkage	0.46 (0.50)	0.45 (0.50)	0.50 (0.50)	0.43 (0.50)	0.44 (0.50)	0.47 (0.50)	0.45 (0.50)	0.41 (0.49)	0.52 (0.50)
Student Clubs Linkage	0.50 (0.50)	0.55 (0.50)	0.54 (0.50)	0.47 (0.50)	0.46 (0.50)	0.51 (0.50)	0.50 (0.50)	0.46 (0.50)	0.57 (0.50)
N	519	99	115	113	119	191	328	302	217

Notes: * denotes significance at 10% level, ** denotes significance at 5% level and *** denotes significance at 1% level where base condition is control. Standard deviations are in parenthesis.

Table 3: Baseline Trust

	Baseline Trust						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Student	Non-Student	Male	Female	White	Minority
Male	-0.765 (2.122)	-1.173 (2.765)	0.140 (3.389)	0.000 (.)	0.000 (.)	-0.835 (2.944)	-0.628 (3.071)
Student	0.771 (2.233)	0.000 (.)	0.000 (.)	0.290 (3.705)	1.053 (2.959)	2.324 (2.922)	-3.453 (3.648)
Age	0.168 (0.129)	0.235 (0.192)	0.189 (0.177)	0.366** (0.174)	0.043 (0.176)	0.065 (0.142)	0.738*** (0.280)
GCSE	Baseline Level						
A-Levels	-2.264 (3.081)	-4.474 (3.700)	0.936 (5.563)	1.068 (5.418)	-4.255 (3.796)	-3.163 (3.776)	-1.545 (5.336)
Undergraduate	6.014** (2.513)	4.407 (3.119)	8.461** (4.251)	8.222* (4.170)	3.974 (3.113)	4.737 (3.248)	7.091* (3.940)
Postgraduate	-2.176 (3.674)	-2.156 (4.541)	-3.401 (6.076)	4.082 (5.462)	-5.712 (4.916)	-4.692 (4.674)	0.853 (5.977)
Missing	-20.057*** (7.375)	-23.144** (9.031)	-18.006 (12.735)	-25.038** (11.268)	-16.375* (8.718)	-15.981 (10.793)	-25.593*** (8.031)
White	Baseline Level						
Black	-7.041** (3.108)	-10.093*** (3.755)	3.934 (4.039)	-1.557 (4.097)	-11.936*** (4.586)		-2.423 (4.617)
Asian	-1.147 (2.806)	-2.530 (3.363)	1.178 (5.443)	-1.886 (4.408)	0.169 (3.924)		2.619 (4.352)
Other	-2.569 (3.739)	-4.432 (4.265)	0.949 (8.001)	-2.427 (6.658)	-2.331 (4.523)		0.000 (.)
Constant	71.138*** (5.183)	72.634*** (6.094)	67.548*** (7.686)	62.208*** (7.885)	76.316*** (6.977)	74.218*** (6.106)	55.236*** (10.464)
Observations	446	287	159	171	275	256	190
R^2	0.059	0.071	0.069	0.086	0.067	0.037	0.118

Notes: * denotes significance at 10% level, ** denotes significance at 5% level and *** denotes significance at 1% level. Standard errors are in parenthesis.

Table 4: Treatment Effect on Trust

	Trust								
	(1) All	(2) All	(3) All	(4) Student	(5) Non-Student	(6) Male	(7) Female	(8) White	(9) Minority
Info	3.782 (2.780)	2.831 (1.807)	2.920 (1.816)	3.969 (2.513)	-0.678 (2.568)	-1.575 (3.642)	5.029** (1.981)	-0.422 (2.222)	8.654*** (2.927)
Info + Photo	2.168 (2.654)	3.573** (1.569)	3.552** (1.596)	3.839* (2.223)	1.108 (2.009)	-1.700 (2.779)	6.718*** (1.956)	2.734 (1.973)	4.722* (2.674)
Photo	0.579 (2.802)	0.691 (1.727)	0.627 (1.796)	-0.507 (2.498)	2.464 (2.166)	0.012 (2.950)	0.653 (2.331)	0.523 (2.489)	1.332 (2.850)
Baseline		0.718*** (0.040)	0.721*** (0.041)	0.658*** (0.058)	0.847*** (0.047)	0.703*** (0.062)	0.752*** (0.050)	0.789*** (0.050)	0.628*** (0.064)
Constant	76.505*** (1.956)	22.486*** (3.344)	20.517*** (4.174)	25.027*** (5.514)	13.533** (5.541)	25.071*** (7.406)	18.004*** (4.988)	14.549*** (5.065)	28.554*** (7.506)
Observations	446	446	446	287	159	171	275	256	190
R^2	0.005	0.589	0.596	0.524	0.771	0.582	0.624	0.631	0.590

Notes: Columns 3-9 control for gender, whether the participant is a student, maternal education level and ethnicity * denotes significance at 10% level, ** denotes significance at 5% level and *** denotes significance at 1% level. Standard errors are in parenthesis.

Table 5: Consent to Linkage

	(1)	(2)	(3)
	Education Records	Health Records	Student Activity Records
Baseline Trust	0.004*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
Male	-0.012 (0.048)	0.003 (0.047)	0.005 (0.047)
Student	0.102** (0.052)	0.000 (0.052)	0.131** (0.052)
Age	-0.004 (0.003)	-0.006*** (0.002)	-0.006** (0.002)
GCSE	Baseline Level		
A-Levels	0.039 (0.065)	0.029 (0.065)	0.027 (0.064)
Undergraduate	0.034 (0.064)	0.050 (0.064)	-0.086 (0.062)
Postgraduate	0.110 (0.080)	0.043 (0.082)	0.024 (0.081)
Missing	-0.190 (0.116)	-0.176* (0.093)	-0.249** (0.099)
White	Baseline Level		
Black	0.192*** (0.070)	0.210*** (0.074)	0.259*** (0.070)
Asian	-0.054 (0.065)	0.046 (0.065)	-0.008 (0.066)
Other	-0.148* (0.075)	0.031 (0.080)	0.041 (0.075)
Constant	0.256* (0.139)	0.221* (0.130)	0.216* (0.129)
Observations	441	441	441
R^2	0.107	0.098	0.133

Notes: * denotes significance at 10% level, ** denotes significance at 5% level and *** denotes significance at 1% level. Standard errors are in parenthesis.

Table 6: TE on Consent to Linkage

	Education Records			Health Records			Student Activity Records		
	(1) All	(2) Student	(3) Minority	(4) All	(5) Student	(6) Minority	(7) All	(8) Student	(9) Minority
Info	0.040 (0.065)	0.021 (0.084)	0.074 (0.104)	0.058 (0.067)	0.036 (0.088)	0.021 (0.109)	-0.006 (0.064)	-0.016 (0.083)	0.003 (0.105)
Info + Photo	-0.039 (0.067)	-0.000 (0.084)	-0.023 (0.103)	-0.006 (0.067)	0.021 (0.084)	-0.017 (0.104)	-0.069 (0.065)	-0.041 (0.082)	-0.051 (0.099)
Photo	-0.021 (0.067)	-0.064 (0.086)	-0.074 (0.104)	-0.007 (0.067)	-0.069 (0.085)	-0.057 (0.104)	-0.094 (0.066)	-0.134 (0.083)	-0.181* (0.101)
Baseline	0.004*** (0.001)	0.004** (0.001)	0.004** (0.002)	0.005*** (0.001)	0.005*** (0.001)	0.004** (0.002)	0.005*** (0.001)	0.005*** (0.001)	0.004*** (0.002)
Observations	441	284	187	441	284	187	441	284	187
R^2	0.110	0.089	0.149	0.101	0.100	0.124	0.140	0.111	0.161

Notes: Regressions control for baseline trust, gender, whether the participant is a student, age, maternal education level and ethnicity. * denotes significance at 10% level, ** denotes significance at 5% level and *** denotes significance at 1% level. Standard errors are in parenthesis.

Appendix A: Additional Tables

Table A1: Treatment Effect on Education Records Linkage

	Education Records						
	(1) All	(2) Student	(3) Non-Student	(4) Male	(5) Female	(6) White	(7) Minority
Info	0.040 (0.065)	0.021 (0.084)	0.093 (0.114)	-0.002 (0.110)	0.060 (0.083)	0.029 (0.085)	0.074 (0.104)
Info + Photo	-0.039 (0.067)	-0.000 (0.084)	-0.102 (0.111)	-0.048 (0.109)	-0.037 (0.089)	-0.062 (0.089)	-0.023 (0.103)
Photo	-0.021 (0.067)	-0.064 (0.086)	0.084 (0.110)	-0.023 (0.104)	-0.019 (0.090)	0.031 (0.089)	-0.074 (0.104)
Baseline	0.004*** (0.001)	0.004** (0.001)	0.005*** (0.001)	0.004* (0.002)	0.004*** (0.001)	0.003** (0.001)	0.004** (0.002)
Observations	441	284	157	169	272	254	187
R^2	0.110	0.089	0.180	0.136	0.105	0.102	0.149

Notes: Regressions control for baseline trust, gender, whether the participant is a student, age, maternal education level and ethnicity. * denotes significance at 10% level, ** denotes significance at 5% level and *** denotes significance at 1% level. Standard errors are in parenthesis.

Table A2: Treatment Effect on Health Records Linkage

	Health Records						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Student	Non-Student	Male	Female	White	Minority
Info	0.058 (0.067)	0.036 (0.088)	0.067 (0.111)	-0.045 (0.114)	0.060 (0.083)	0.088 (0.086)	0.074 (0.104)
Info + Photo	-0.006 (0.067)	0.021 (0.084)	-0.073 (0.117)	-0.022 (0.105)	-0.037 (0.089)	-0.009 (0.088)	-0.023 (0.103)
Photo	-0.007 (0.067)	-0.069 (0.085)	0.095 (0.113)	0.026 (0.105)	-0.019 (0.090)	0.033 (0.090)	-0.074 (0.104)
Baseline	0.005*** (0.001)	0.005*** (0.001)	0.004*** (0.002)	0.006*** (0.002)	0.004*** (0.001)	0.004*** (0.001)	0.004** (0.002)
Observations	441	284	157	169	272	254	187
R^2	0.101	0.100	0.132	0.179	0.105	0.092	0.149

Notes: Regressions control for baseline trust, gender, whether the participant is a student, age, maternal education level and ethnicity. * denotes significance at 10% level, ** denotes significance at 5% level and *** denotes significance at 1% level. Standard errors are in parenthesis.

Table A3: Treatment Effect on Student Activities Records Linkage

	Student Activity Records						
	(1) All	(2) Student	(3) Non-Student	(4) Male	(5) Female	(6) White	(7) Minority
Info	-0.006 (0.064)	-0.016 (0.083)	0.016 (0.108)	-0.049 (0.114)	0.011 (0.081)	-0.002 (0.085)	0.003 (0.105)
Info + Photo	-0.069 (0.065)	-0.041 (0.082)	-0.117 (0.107)	-0.087 (0.110)	-0.064 (0.087)	-0.089 (0.089)	-0.051 (0.099)
Photo	-0.094 (0.066)	-0.134 (0.083)	-0.012 (0.106)	-0.146 (0.102)	-0.060 (0.090)	-0.023 (0.088)	-0.181* (0.101)
Baseline	0.005*** (0.001)	0.005*** (0.001)	0.005*** (0.001)	0.006*** (0.002)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.002)
Observations	441	284	157	169	272	254	187
R^2	0.140	0.111	0.186	0.231	0.103	0.127	0.161

Notes: Regressions control for baseline trust, gender, whether the participant is a student, age, maternal education level and ethnicity. * denotes significance at 10% level, ** denotes significance at 5% level and *** denotes significance at 1% level. Standard errors are in parenthesis.

Appendix B: Screenshots from Treatment

Figure B1: Screenshots from the Treatment Arms

B1.1: Information Treatment

The lead researcher in this project is Dr Joshua Fullard.

Joshua Fullard is a lecturer in the Department of Economics at the University of Essex. Joshua received his PhD from the Institute for Social and Economic Research and has previously worked as a senior researcher at the Education Policy Institute and as a visiting research fellow at the ifo Institute in Munich. His work has gained national and international media coverage in a range of publications including: the Times, Telegraph, Guardian, Economist, Daily Mail and China Daily.

B1.2: Photo Treatment

The lead researcher in this project is Dr Joshua Fullard.



B1.3: Combined Treatment

The lead researcher in this project is Dr Joshua Fullard.



Joshua Fullard is a lecturer in the Department of Economics at the University of Essex. Joshua received his PhD from the Institute for Social and Economic Research and has previously worked as a senior researcher at the Education Policy Institute and as a visiting research fellow at the ifo Institute in Munich. His work has gained national and international media coverage in a range of publications including: the Times, Telegraph, Guardian, Economist, Daily Mail and China Daily.