HACETTEPE UNIVERSITY INSTITUTE
OF POPULATION STUDIES

IDENTIFYING THE CONSENT BEHAVIOR TO LINK SURVEY AND ADMINISTRATIVE DATA

Türknur BRAND

Department of Social Research Methodology

PhD Thesis

Ankara
June 2024

HACETTEPE UNIVERSITY INSTITUTE
OF POPULATION STUDIES

IDENTIFYING THE CONSENT BEHAVIOR TO LINK SURVEY AND ADMINISTRATIVE DATA

Türknur BRAND

Supervisors: Prof. Dr. Ahmet Sinan TÜRKYILMAZ
Prof. Dr. Joseph W. SAKSHAUG

Department of Social Research Methodology
PhD Thesis

Ankara
June 2024

APPROVAL PAGE

This is to certify that we have read and examined this thesis and, in our opinion, it fulfills the requirements in scope and quality of a thesis for the degree of Doctor of Philosophy in Social Research Methodology.

Jury Members:

Member (Chair): ……....
Prof. Dr. Berna Burçak BASBUG ERKAN Middle East Technical University, Faculty of Arts and Sciences, Department of Statistics

Member (Supervisor): ……....
Prof. Dr. A. Sinan TÜRKYILMAZ Hacettepe University, Institute of Population Studies, Department of Social Research Methodology

Member (Co-supervisor): ……....
Prof. Dr. Joseph W. SAKSHAUG, Institute for Employment Research, Ludwig Maximillian University of Munich

Member: ……..
Prof. Dr. Yaprak Arzu ÖZDEMİR Gazi University, Faculty of Sciences, Department of Statistics

Member: ……....
Assoc. Prof. Dr. İlknur YÜKSEL KAPTANOĞLU Hacettepe University, Institute of Population Studies, Department of Social Research Methodology

Member: ……....
Assoc. Prof. Dr. Ayse ABBASOGLU ÖZGÖREN Hacettepe University, Institute of Population Studies, Department of Demography

Member: ……....
Asst. Prof. Dr. Melike SARAC Hacettepe University, Institute of Population Studies, Department of Social Research Methodology

This thesis has been accepted by the above-signed members of the Jury and has been confirmed by the Administrative Board of the Institute of Population Studies, Hacettepe University.

……....……. …/…/ 2024

Prof. Dr. İsmet KOÇ
Director

HACETTEPE UNIVERSITY INSTITUTE OF POPULATION STUDIES
THESIS ORIGINALITY REPORT

HACETTEPE UNIVERSITY
INSTITUTE OF POPULATION STUDIES
THESIS/DISSERTATION ORIGINALITY REPORT

HACETTEPE UNIVERSITY
INSTITUTE OF POPULATION STUDIES
TO THE DEPARTMENT OF SOCIAL RESEARCH METHODS

Date: 12/08/2024

Thesis Title / Topic: **Identifying the Consent Behavior to Link Survey and Administrative Data**

According to the originality report obtained by myself/my thesis advisor by using the Turnitin plagiarism detection software and by applying the filtering options stated below on 04/08/2024 for the total of 108 pages including the a) Title Page, b) Introduction, c) Main Chapters, and d) Conclusion sections of my thesis entitled as above, the similarity index of my thesis is 5 %.

Filtering options applied:

1. Bibliography/Works Cited excluded
2. Quotes excluded
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Institute of Population Studies Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

12/08/2024

Name Surname:	Türknur Brand
Student No:	N16149396
Department:	Social Research Methodology
Program:	Social Research Methodology- Doctorate
Status:	☐ Masters ☒ Ph.D. ☐ Integrated Ph.D.

ADVISOR APPROVAL

APPROVED.

Prof. Dr. Ahmet Sinan Türkyılmaz

12/08/2024

Turknur Brand_PhDThesis_04082024_Turnitin.docx

by Ahmet Sinan Türkyılmaz

Submission date: 04-Aug-2024 08:30PM (UTC+0300)
Submission ID: 2427107258
File name: Turknur_Brand_PhDThesis_04082024_Turnitin.docx (984K)
Word count: 23647
Character count: 129749

Turknur Brand_PhDThesis_04082024_Turnitin.docx

ORIGINALITY REPORT

5%	4%	4%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	madoc.bib.uni-mannheim.de Internet Source	1%
2	academic.oup.com Internet Source	1%
3	www.studystack.com Internet Source	1%
4	"Big Data Meets Survey Science", Wiley, 2020 Publication	1%
5	repository.essex.ac.uk Internet Source	<1%
6	core.ac.uk Internet Source	<1%
7	digitalcommons.library.uab.edu Internet Source	<1%

ETHICAL DECLARATION

In this thesis study, I declare that all the information and documents have been obtained on the basis of the academic rules and all audio-visual and written information and results have been presented according to the rules of scientific ethics. I did not do any distortion in data set. In case of using other works, related studies have been fully cited in accordance with scientific standards. I also declare that my thesis study is original except the cited references. It was produced by myself in consultation with my supervisors (Prof. Dr. A. Sinan TÜRKYILMAZ, Prof. Dr. Joseph W. SAKSHAUG) and written according to the rules of thesis writing of Hacettepe University Institute of Population Studies.

…….
Türknur BRAND

DECLARATION OF PUBLISHING AND INTELLECTUAL PROPERTY RIGHTS

I declare that I give permission to Hacettepe University to archive all or some part of my PhD thesis, which is approved by the Institute, in printed (paper) or electronic format and to open to access with the following rules. With this permission, I hold all intellectual property rights, except using rights given to the University, and the rights of use of all or some parts of my thesis in the future studies (article, book, license, and patent).

I declare that the thesis is my original work, I did not violate rights of others and I own all rights of my thesis. I declare that I used texts with the written permit which is taken by owners and I will give copies of these to the University, if needed.

As per the "Regulation on the Online Availability, Arrangement and Open Access of Graduate Theses" of Council of Higher Education, my thesis shall be deposited to National Theses Center of the Council of Higher Education/Open Access System of H.U. libraries, except for the conditions indicated below;

- The access to my thesis has been postponed for 2 years after my graduation as per the decision of the Institute/University board. (1)
- The access to my thesis has been postponed for …. month(s) after my graduation as per the decision of the Institute/University board. (2)
- There is a confidentiality order for my thesis. (3)

09/08/2024

…….
Türknur BRAND

---

(1) Article 6.1. In the event of patent application or ongoing patent application, the Institute or the University Board may decide to postpone the open access of the thesis for two years, upon the proposal of the advisor and the assent of the Institute Department.

(2) Article 6.2. For theses that include new techniques, material and methods, that are not yet published articles and are not protected by patent and that can lead to unfair profit of the third parties in the event of being disseminated online, the open access of the theses may be postponed for a period not longer than 6 months, as per the decision of the Institute or the University Board upon the proposal of the advisor and the assent of the Institute Department.

(3) Article 7.1. The confidentiality order regarding the theses that concern national interest or security, the police, intelligence, defense and security, health and similar shall be issued by the institution certified the thesis*. The confidentiality order for theses prepared pursuant to the cooperation protocol with institutions and organizations shall be issued by the University Board, upon the proposal of the related institutions and organizations and the assent of the Institute or the Faculty. The theses with confidentiality order shall be notified to the Council of Higher Education.

Article 7.2. During the confidentiality period, the theses with confidentiality order shall be kept by the Institute or the Faculty in accordance with the confidentiality order requirements, in the event of termination of the confidentiality order the thesis shall be uploaded to Thesis Automation System.

* Shall be issued by the Institute or Faculty Board upon the proposal of the advisor and the assent of the Institute Department

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Ahmet Sinan Türkyılmaz for his support and encouragement.

I would like to thank Professor Joe Sakshaug for being a second dissertation advisor and supporting me with his advice and guidance on this topic.

I greatly appreciate the time and effort that my dissertation committee gives in all circumstances.

I would like to thank my dear friends Songül Cömert, Ömür Kayıkçı Çiçek, Aydan Golaszewski, Leanore Regensburger, Rashna Ghadialy, Selen Başer, Nazan Alkan. They were always with me; listening me, encouraging me, supporting me, and giving me strength, no matter what part of the world I was in. This thesis made three countries; Türkiye, Germany, United States.

I would also like to thank my parents (especially my dad, Mehmet Hamsici), my brother, my husband, and his parents (especially my mother-in-law, Ingrid Brand) for their support, encouragement, and understanding throughout this Ph.D. journey.

Finally, I would like to thank my dear son Friedrich Ilker for his patience and love at all stages. He deserves the greatest applause.

ABSTRACT

This thesis aims to understand the determinants of consent behavior in linking survey and administrative data, from two perspectives; one with the metadata formed using the surveys with consent questions for data linkage, namely the macro approach and the other perspective is the analysis of the General Social Survey (GSS) which is a survey with consent for data linkage, namely the micro approach. There are two main objectives; the first one is to find the significant survey characteristics to influence the consent decision and similarly, the second one is the consent asking procedures to influence the consent behavior. Given this objective, in the first stage, a systematic review is conducted to analyze and present the results of survey research that uses consent for data linkage in their survey. At this stage, there is a wide variety of surveys from different countries that constitute the data of this part. The systematic review procedures are followed to collect these data. Through systematic review statistical procedures, the study evaluates numerous factors of survey and consent design characteristics with constructed variables such as survey response rate, topic, country, year, type, mode, age of target population, sponsor, and various aspects related to the consent request process. The continuous consent rate is modeled with multiple and stepwise approaches. The second section examines consent research at the micro-survey level, using the GSS as an example. The consent rate, a binary dependent variable, is explained by binary logistic regression. The results of both perspectives indicate that respondents' trust in organizations and trust, in general, are related to their consent decisions. The other identified variables seem to have less impact on consent rates in these surveys.

Keywords: Data collection, survey methods, semiparametric and nonparametric methods

ÖZET

Bu tezin amacı, anket ve idari verileri birleştirilme aşamasında verdikleri onay davranışının belirleyicilerini iki açıdan anlamaktır; birincisi veri birleştirilmesi için onay soruları içeren anketler kullanılarak sistematik olarak derlenmiş verilerden oluşmaktadır, ikincisi ise veri birleştirilmesi için onay alma sorusunu içeren General Social Survey (GSS) analiz edilmektedir. Bunlar yapılırken iki temel amacımız vardır; birinci amaç, anket katılımcılarının onay kararı almasını etkileyecek önemli anket özelliklerini bulmak ve benzer şekilde ikincisi ise, onay verme davranışını etkileyebilecek olan onay isteme prosedürleridir. Bu hedefler dikkate alındığında, birinci yöntem, anketlerinde veri birleştirilmesini gerçekleştirmek için onay soran anket araştırmalarını analiz etmek ve sunmak için sistematik bir inceleme yapılmıştır. Bu çalışmada birçok ülkeden çok çeşitli anket verileri bulunmaktadır. Bu verileri toplamak için sistematik derleme adımları takip edilmiştir. Çalışmada sistematik derleme yöntemi kapsamında istatistiklerle, anket cevaplama oranı, anket konusu, anket yılı, anket veri toplama metodu, anket sponsoru ve anket veri birleştirme onay isteme metodu gibi oluşturulmuş değişkenlerle anket ve onay tasarımı özelliklerine ilişkin etkenler değerlendirilmektedir. Bu değişkenler kullanılarak onay verme davranışı, regresyonlarla modellenmiştir. İkinci aşamada ise onay verme davranışı araştırmaları, mikro anket verisi olan General Social Survey (GSS) örneğiyle incelenmektedir. Her iki aşamada da regresyon modelleri, onay verme davranışı alt değişkenlerini keşfetmek için kullanılmaktadır. Her iki araştırmanın da sonuçları, katılımcıların kuruluşlara olan güveninin ve genel olarak güvenin, onay verme kararlarıyla ilişkili olduğunu göstermektedir. Belirlenen diğer değişkenlerin bu anketlerdeki onay verme davranışı oranları üzerinde daha az etkisi olduğu görülmektedir.

Anahtar Kelimeler: Veri toplama, anket metotları, yarı parametrik ve parametrik olmayan metotlar

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ANOVA	Analysis of Variance
CAPI	Computer-Assisted Personal Interviewing
CATI	Computer-Assisted Telephone Interviewing
CDS	Computerized Delivery Sequence
EPSEM	Equal Probability Selection Method
GDPR	General Data Protection Regulation
GSS	General Social Survey
IAP	Inapplicable
IRB	Institutional Review Board
ISSP	International Social Survey Program
MSA	Master Shipping Account
TSE	Total Survey Error
TURKSTAT	Turkish Statistical Institute
USPS	United States Postal Service
VIF	Variance Inflation Factor

CHAPTER 1. INTRODUCTION

Consent, a commonly used word, means a voluntary agreement by a person or institution to the wish or proposal of another person or institution. This thesis aims to investigate consent in the context of survey research, specifically data linkage consent. Administrative data is another data source collected for public or commercial reasons, such as campaigns, government benefits, etc., or that may contain some general data like demographics or specific information. Identity information and address are tools used to link survey and administrative data. These linkages have been used for many years (Blumberg & Cynamon, 1999). In addition, these linkages are also possible for administrative records to administrative records or surveys to another survey. Lately, the technological capabilities of the new era facilitate all processes. Thus, it has been much easier to link surveys and the other data sources mentioned As Smith & Kim, (2013) explain new methods helped to link surveys and other data much more efficiently. This research stemmed from a curiosity about the so-called consent of survey respondents to link their data with administrative records.

In the survey methodology literature, there are three types of consent. The first is the widely known informed consent to participate in a survey. Accepting to participate in a survey is a free choice of the person being sampled. Since the "National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research" (1979) declared that respondents have the right to be protected from harm and to decide whether to participate or not to participate in research, informed consent has been discussed. This means maintaining the confidentiality of the data collected by the researchers and obtaining the informed consent of the respondents. The use of various types of surveys raises ethical and practical questions about the application of informed consent. Singer, (1984) identified concerns about respondent confidentiality in social research on attitudes and behaviors. She has studied all aspects of informed consent and published research on ethical issues for many years, from 1978 to 2017. She found the importance of confidentiality, both ethically and practically, and mentions the need for adequate

protection by both researchers and committees reviewing informed consent. She assessed the poor understanding of informed consent in both social and biomedical research. She clarified its meaning through many experiments and by using the method for many years. She also examined ways to document consent, with or without a signature, using methods of experimental research design. In addition, the survey literature discusses many methods used in informed consent to motivate survey participation. Incentives in the form of gift cards, cash, etc. are widely accepted. Depending on the sample, thank you letters may be preferred to cash gifts as a token of appreciation.

The second type of consent is to obtain permission to take some measurements, such as a blood test or other physical measurements, usually in health surveys. Survey participants may or may not want to share physical health measures of their own free will. This type of consent also has legal and procedural confidentiality protections known to researchers and review boards. There is a higher number of positive responses to consent questions in the older group (Jäckle et al., 2021). It is assumed that the interest shown in their health data by the survey organizations could mean a closer look and solution to their health problems.

A third type of consent is sought by survey agencies for data linkage to respondents' administrative records. This third type of consent is the subject of this paper. Below are two examples of consent requests in two surveys focused on health and business (Fulton, 2012):

"Health-Related Administrative Record Consent Request:

We would like to understand how the use of health care may change as people age. To do that, we need to obtain information about vital statistics, health care costs, and diagnoses from your health-related records. In order for us to retrieve these records, we need your consent. This will allow us to conduct more research without asking additional questions. Your consent is voluntary and the information that you provide will be kept completely confidential. May I have your consent to access these records?"

"Income and Employment-Related Administrative Record Consent Request:

We would like to understand how people's income changes as they age. To do that, we need to obtain information about income and employment from your income and employment-related records. In order for us to retrieve these records, we need your consent. This will allow us to conduct more research without asking additional questions. Your consent is voluntary and the information that you provide will be kept completely confidential. May I have your consent to access these records?"

In the last century, the interest in accessing as much data as possible in all areas of life has become very popular, resulting in a huge demand for data research (Kim, J.K. & Rao, J.N., (2012); Fobia, A.C. et al., (2019)). Survey data and administrative records are linked in the case of two main necessities. One is to fill in missing cells in the dataset for various reasons specific to the survey design. If the administrative and survey data can be linked using the correct identifier information, missing survey data can be completed. The second is the time advantage, since not all information about respondents, usually demographic information, is necessarily collected during the survey. The use of administrative data to collect complete population records can lead to shortened sample questionnaires, thereby reducing time, cost, and respondent burden. Any survey may pursue one or both of these objectives within its design framework. Thus, the combination of survey and administrative records is intended to increase the amount of information available in each unit and to support analysis for research purposes. Government agencies and other organizations use this tool to collect administrative data for registration, transaction, and record-keeping purposes, while survey data are based on a more targeted population and include specific questions. In this way, data linkage is widely used where it is feasible. Besides, the most significant point that matters to the survey research is the decreased errors or bias using data linkage.

Although linking administrative and survey data is simple in theory, the practical implementation of finding all matching records by name, date of birth, or other identifying information has proven to be extremely complex. This is especially true when trying to

link millions of population records. It is well known that over time, increasing survey costs and declining survey participation have led to increased demand for data, but tighter budgets. As a result, it has been advantageous to link data that is partly survey data and partly administrative data.

Expectedly there are legal procedures that provide guidelines on how these steps of linking data to registers can be done. As it is known, the privacy and confidentiality of certain data are important to people and institutions. There is a key variable for data linkages to merge survey data with administrative data. In most situations, this variable is some kind of identification number, such as a social security number, a health registry card, an identity card, or something similar that is private and personal. Therefore, the sharing of sensitive data with any authorities is a critical debate to build strong structural steps. Legal authorities have described laws and ethical standards on how the researchers and institutions can proceed to do linkage. For example, data linkage is discussed under the Institutional Review Board (IRB) in the US, an organization that reviews research studies to ensure that they comply with applicable regulations, meet generally accepted ethical standards, follow institutional policies, and adequately protect research participants. IRBs protect the rights and welfare of human subjects in research activities.

Similarly, in times of increasing interest in big data, "General Data Protection Regulation (GDPR)" in EU area, a law was passed on April 27, 2016, with the implementation date after May 25, 2018. This law mainly takes into account the privacy and confidentiality of any type of personal data. Consent in the GDPR law was identified as follows:

"*Article 7 Conditions for consent*

1. *Where processing is based on consent, the controller shall be able to demonstrate that the* ***data subject has consented to processing of his or her personal data****.*

2. If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a

manner which is clearly distinguishable from the other matters, ***in an intelligible and easily accessible form, using clear and plain language****. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.*

3. ***The data subject shall have the right to withdraw his or her consent at any time****. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof.* ***It shall be as easy to withdraw as to give consent.***

4. When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

11) 'consent*' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;*

1. Processing shall be lawful only if and to the extent that at least one of the following applies:

(a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

(32) Consent *should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the* ***data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement.*** *This could include ticking a box when visiting an internet website, choosing technical settings for information society services or another statement or conduct which clearly indicates in this context the data subject's acceptance of the proposed processing of his or her personal data. Silence, pre-ticked boxes or inactivity should not*

therefore constitute consent. Consent should cover all processing activities carried out for the same purpose or purposes. ***When the processing has multiple purposes, consent should be given for all of them.*** *If the data subject's consent is to be given following a request by electronic means, the request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided."*

Citation: « GDPR: regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec ()»

The GDPR law has been a game-changer for many institutions in the EU in the follow-up in many countries, including the US. From now on, the mandatory initial step in creating a link between the survey and registered data is to obtain permission from participants to use their private and confidential information for any purpose. Such permission can result in either consent or refusal. Also, as emphasized in the GDPR, respondents can change their minds and refuse to give consent anymore at any time. The language of the consent question should be very precise and easy to understand. In a survey round, the proportion of respondents who give consent to link their data to all survey respondents is the calculation of the consent rate for data linkage.

In this thesis, we intend to understand the situations that result in a "yes I consent" or "no I do not consent" response for the consent to data linkage. As a part of the first research, we intend to reach out to as many surveys as possible asking consent for data linkage under a method of systematic review guidelines which includes reading some manuscripts and making decisions to include data or not under certain inclusion and exclusion criteria. Data comes from those surveys that fit the criteria. Then the possible components of consent behavior were discussed to create variables and work on them with analyses. This constitutes the macro part of the study. As a next step, having learned from

the first part, macro perspective analyses of the thesis, one of the survey studies GSS data in 2021 is analyzed for expected identifiers at a micro level.

With all that, the objective is to explain consent behavior at least to a certain extent, and seek to answer the following research questions:

(1) Which survey design characteristics significantly impact the rate of consent?
(2) Which consent design characteristics significantly impact the rate of consent?

There are seven chapters in this thesis. The initial chapter is the introduction providing information on key issues of the thesis; and consent for data linkage from a more general perspective. The concept of consent types is introduced at a universal level and the type of consent of interest to this thesis is explained briefly. The second chapter reviews the literature for the selected group of research on the characteristics identified to be influential for consent behavior. Theories relevant to the subject and theoretical framework are discussed in this chapter. The third section explains the methodology including research strategy, inclusion-exclusion criteria, survey selection, data set, and preparation of variables for statistical analysis. The fourth chapter analyzes the General Social Survey as one example to witness what determinants make a difference in consent behavior. In the fifth section, the results of the first macro approach with systematic review are evaluated. In the sixth section, findings of the micro approach on GSS data are assessed. Afterward, all descriptive and regression analyses implemented are discussed. In the final section, results, limitations, present implications for further work on consent rates, and recommendations on how to structure consent to data linkage questions in surveys are discussed.

CHAPTER 2. LITERATURE AND THEORETICAL FRAMEWORK

2.1. Literature

Since the beginning of survey and administrative data linkage, there has been an ongoing debate in the literature about who gives consent and who does not. These questions are similar to those raised in the literature also questioned for the first type of consent, on informed consent. Although the research community has conducted many experiments and studies on consent behavior for data linkage over the past decade, there is still no consensus in the literature explaining the mechanics of survey and consent design (Baghal et al., 2019; Carter et al., 2010; Sakshaug & Kreuter 2014; Sala et al., 2012). Studies in the literature vary in populations, modes, and consent methods, so it is not advisable to directly generalize findings from one study to others (Kreuter et al., 2016).

To date, among many characteristics, the placement and wording of the consent question have been the most frequently observed features. Sakshaug and Vicari (2018) found that in a web survey of establishments, the consent question for linkage placed at the beginning of the survey received a positive response rate of 61.3 percent. Similarly, in a Computer-Assisted Telephone Interviewing (CATI) survey of employment, the consent rate is 95.6 percent when the question is placed at the beginning of the survey (Sakshaug et al.,2013). On the other hand, when the consent question is placed anywhere in the context of a Computer-Assisted Personal Interviewing (CAPI) survey of households, the consent rate is 65 percent (Sala et al., 2014). Although the location of the consent question in the survey is effective, the mode of the survey is also influential. Some of the features are interactive. Thus, the location of the consent question in the survey is not, by itself, a single factor that changes the direction of linkage consent. Many studies like these give researchers a clue that consent behavior is never the result of a single factor; it is a combination of a variety of characteristics.

Jäckle et al. (2021) want to investigate the factors that influence the decision to consent using survey data from the UK in 2018 and 2019. Accordingly, they want to develop some methods to increase the consent rate, especially in web surveys. Therefore,

they are conducting some experiments to observe the different features of the consent request, how they influence different characteristics of respondents, and how the decision works. Another question is to find out how consent for data linkage varies between experimental groups and how respondent characteristics differ between face-to-face and self-administered web surveys. They found that providing more information in the consent question is not necessarily helpful because respondents make a consent decision quickly. They claim that it is better to keep the format as simple as possible and to work on the trust between the survey organization and the respondents. These relationships were found to be more important than anything else in maintaining a higher consent rate. They point to the importance of research to understand how respondents make responsible decisions and increase consent rates, especially in online surveys. The results of this experimental research provide ideas on how to better follow up on online consent requests.

Grove's (2000) Leverage Salience Theory explains differences in survey participation by making the survey more appealing to respondents through methods such as incentives and benefit framing. Since most research on survey participation relies on bivariate analyses, Groves et.al., (2000) wanted to test multiple factors to build this theoretical framework, called Leverage Salience Theory. Leverage salience theory suggests that each person has different leverage on the decision to participate in a survey depending on the given survey attributes. Incentives are popular and are usually gifts or small amounts of money given to respondents as a token of appreciation after they have responded. Benefit framing is another feature that uses a detailed explanation of the purpose of this consent for the benefit of the respondent. One or both of these usually help participants make a decision. Previous research by Groves and Couper (1998) shows that respondents vary in the characteristics of the survey request that they judge to be relevant to their decision. These judgments are influenced by each person's background and experiences in ways unknown to the interviewer (or survey designer) at the time of the request. Leverage is a property of a sample person's evaluation of a particular attribute of a survey request; a $10 incentive is likely to produce different leverage than a $20

incentive for a given sample person, but the direction and magnitude of the leverage may vary across individuals. Groves et al. (2000) use logistic regression models to examine the likelihood of cooperation with the survey as a function of incentives, community involvement, and other sociodemographic covariates such as gender, age, etc. In a multifactor environment, the positive influence of an incentive is diminished when cooperation is motivated by community involvement. For example, in a multi-factor environment, the positive influence of an incentive is reduced when community involvement motivates cooperation. With this research, it is found that evaluating the combined effect of factors works efficiently for informed consent for survey participation, which is likely a similar approach to consent for data linkage. As consent for data linkage is related to the literature on informed consent, this feature was investigated by Sakshaug et al. (2013). Placing relevant linkage consent questions in a beneficial frame can make questions more interesting to respondents. What is said to explain the benefit frame and how the respondent interprets it as beneficial is important. Is the reason for consenting to data linkage time-saving by answering fewer questions, a scientific purpose, linkage helping with an unknown answer, or something else?

For busy respondents, time-saving information is more effective than neutral questions in obtaining higher consent rates in a CATI survey of employees (Sakshaug et al., 2019). However, in a CATI survey of US voters, negative wording such as "less useful" that emphasizes the loss received 66.8% consent (Kreuter et al., 2016). At some point, it is unpredictable, human behavior, it depends on the respondent's perception of their role in life. According to Elevelt's (2021) review, the survey sponsor and the wording are the most important identifiers of consent behavior.

In Türkiye, since the beginning of 2018, the Turkish Statistical Institute (TURKSTAT), a governmental organization, has started to use administrative data for its survey purposes in the production of a part of business statistics, namely construction statistics. However, this linkage is implicitly applied and no consent is asked to the respondents of the business surveys. There is a different data protection law in Türkiye,

that is the KVKK and the Board which establishes regulations and oversees the provisions. The KVKK serves a mostly administrative and government-relations role, whereas the Board is the decision-making organ within the authority. The Board began operating in January 2017, once all appointments were made. The Board comprises nine members, elected as follows; five elected by the National Grand Assembly of Türkiye and four directly appointed by the President of Türkiye. Government authorities are the main figures for the data protection role in Türkiye.

2.2. Theoretical Framework

Although, technically, the consent for data linkage framework seems to have similarities in terms of error and bias formation as informed consent for survey participation, a respondent's refusal or agreement to participate in the survey has more components compared to consent for data linkage. Psychologists and sociologists have studied survey response behavior for many years. Tourangeau et al. (2000) in their book explain this behavior as follows; at the stage of answering any type of question, the first step is to understand the question by using memory to retrieve relevant information, and then prepare the evaluation in the required format. And in the process of taking these steps, respondents are assumed to be opportunists or pragmatists. Cognitive scientists are more focused on the causes of errors while statisticians question the consequences of different types of survey errors on the estimates derived from the survey. Both causes of errors and consequences of errors are equally important. From a statistical perspective, survey errors have two main consequences for survey estimates. If the errors are systematic, the estimate is biased; if the errors are random, the variance is increased.

As known, the goal of any survey is to make a generalization about a target population. With this in mind, the best procedures are used to select a representative sample with minimal measurement error. Any type of missingness is to be avoided during data collection for a designed survey. Missingness causes different types of errors for

different estimates about the population which will be explained further and can be seen in Figure 1.

Over the last century, survey data has become an important source of data collection methods for researchers in many fields, including political science, social science, economics, and history. Survey methodologies vary, and results can be accurate or inaccurate, so predictions are subject to error. Therefore, quality is the main issue to draw valid conclusions. For any survey research, the research results must be the most correct with the least errors in all steps from design to field period. Therefore, the so-called sources of error in the entire survey literature explained below are very helpful in identifying and preparing any errors in a disciplined and standardized way in all sciences. The next step is to take the necessary precautions when designing the survey. The ideal approach for any type of survey is to design the best possible survey under the circumstances of the survey environment. These efforts are made before data collection to minimize the possibility of survey error. Working on an ideal design before the field period has a great advantage in reducing costs. During data collection, there may still be actions that can be taken to address problems that are encountered. However, after the data is collected, there are more costs involved in correcting anything, and maintaining quality starts to become a concern. From the beginning of the survey, where a population is targeted, to the end, where data is collected, all phases can cause different types of errors. With the theory of total survey error (TSE) developed in survey methodology, the literature is extremely knowledgeable about the errors that cause bias or errors. Both the design stages and the quality of surveys are discussed under this framework (Figure 2.1.). Bias is a systematic error and variance is a random error. Both can affect the validity of survey data.

As said, this framework covers both the design phases and, as an output, the quality of surveys. The TSE perspective includes the sources of systematic error, i.e. bias, and random error, i.e. variance, that can affect the accuracy of survey data. TSE is also defined as the sum of sampling error and sources of non-sampling error from a measurement

perspective; or the sum of coverage error, sampling error, non-response error and measurement error; or a more modern approach by grouping the various sources of error into the classes of representation (coverage, sampling, non-response, adjustment error) and measurement (validity, measurement error, processing error).

Measurement error, which is a type of non-sampling error, is the difference between the true and observed values. These errors are discussed as a result of interviewer, respondent, and data collection tools errors. For example, inaccuracies or disaccuracies during the data collection may be caused by biased questionnaires, or inexperienced interviewers, which are misinterpreting, or misleading of question intentions (Axel, 2020; Simoen et al., 2015). As a result, the answers are misleading, they are not the true values measured.

Validity is in place when researchers' intentions are satisfied with the measurement tool. The survey item has only validity if the respondent answers in parallel to the intentions of the survey question asked. In fact, questions reflect the intentions of the research correctly. From a cognitive perspective, the respondent understands the intentions of the research with the indicated questionnaire and answers accordingly (Krosnick, 2018).

Processing errors are the mistakes that occur during coding, editing, transcription, data entry, data cleaning, and tabulation. These errors happen after data is collected usually due to technological problems, human mistakes, or inconsistent data processing guidelines. If not found and corrected, survey results will be misinterpreted and lead to false conclusions. Depending on the amount of processing errors, they can introduce biased results as a consequence of validity and reliability concerns of findings. Quality control measures such as automated data checks, double data entry, and data audits are engaged to diminish processing errors. Most importantly, humans’ careful attention to detail, correct documentation, and trustworthy processing tools guarantee coherent research results.

Coverage error, a nonsampling error, occurs when there is not one to one correspondence between the target population and sample frame, that is, some members of a population are excluded from the sample frame used for the study. The degree of coverage error is determined by the population of interest. Commonly seen coverage errors are telephone surveys excluding people without telephones, cell phones, and internet surveys excluding non-internet users or areas with no access to the Internet.

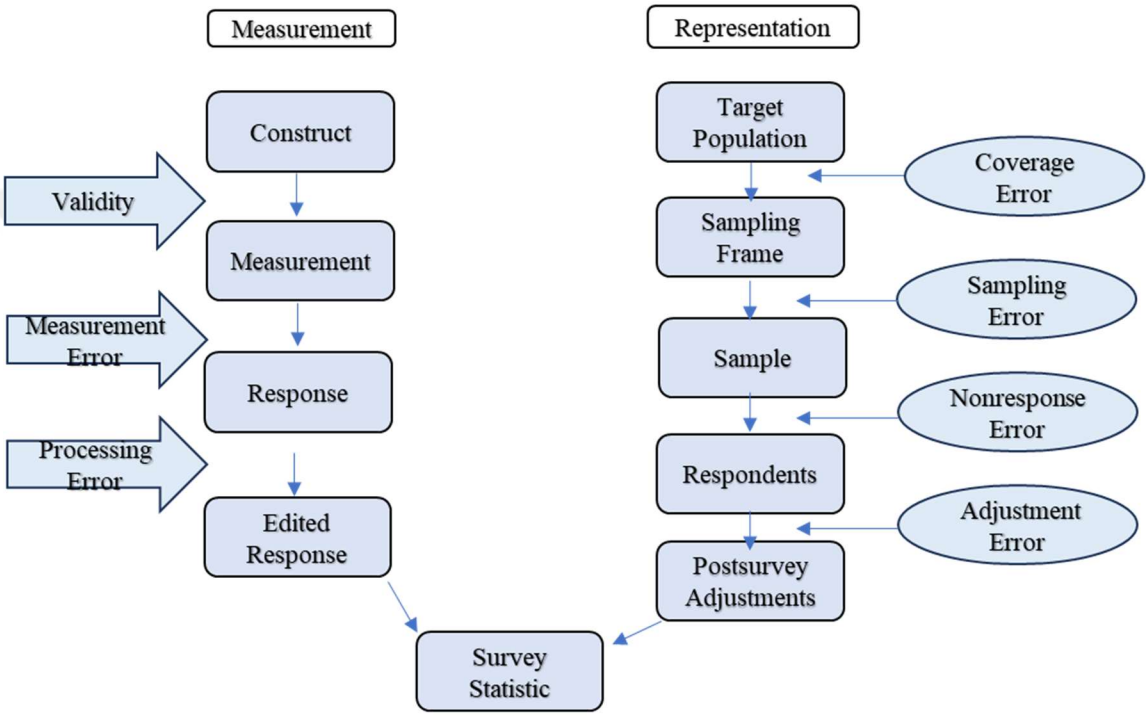
Sampling error, differences between sample construct and the actual value of population. This happens because the sample drawn does not represent the population of interest. If there is a randomly selected sample or a bias, sampling error is calculated.

Nonresponse error occurs when the selected respondents do not respond to the survey. There are two types of nonresponse, item nonresponse and unit nonresponse. Item nonresponse is related to not responding to some of the questions in the survey. Unit nonresponse is not participating in the survey, basically not responding to any of the questions, which is discussed as a nonresponse error. There are reasons for not being available, not willing, and not capability to respond. As in coverage, it is not true to interpret the level of nonresponse rate by itself. For a specific question, it is important to find out if respondents and nonrespondents act systematically differently which causes bias and, as a result, nonresponse error. In other words, the degree of error can fluctuate at the question level of the same research, with the same nonresponse rate.

Adjustment error is calculated as a result of adjustments after data is collected, basically the difference between the adjusted statistic and the population parameter. Those adjustments intend to increase the quality of survey estimates by adjusting coverage, nonresponse, and sampling errors.

Figure 2.1. Total Survey Error Diagram

Source: Groves R.M. et al. (2004)



With the integration of new methods of linking survey data and administrative records, we can speak of new types of error, namely nonconsent and linkage errors, which are to be explained under the representation category of TSE. Some studies have already found important differences in the survey constructs of consenting and non-consenting participants. As a result, some survey estimates may be biased. Since there is usually no information on non-consenters, it is difficult to estimate bias. For example, Sakshaug and Kreuter (2012) used the German Panel Study survey to estimate non-consent biases. For this survey, they had all administrative data on consenters and non-consenters. They found that few estimates have the non-consent biases to a relatively smaller amount compared to other types of biases, such as non-response and measurement.

The theory of TSE explains this for each stage at the measurement and representativeness stages. Why is missingness a major problem in the theory? Why does it interfere with making the best estimates? The whole debate stems from the randomness of missingness. Random missingness means that there is no systematic pattern, it is unobserved and unpredictable. Rubin (1976) states that when making inferences about a population parameter, anything that causes missing data can be ignored if only missing data is randomly missing and observed data is randomly observed. Once missing reasons are not random then there is a systematic difference between sample parameter estimates and population parameters. This is called bias. Multiplying each type of noncoverage, nonresponse, etc. rates, with its bias gives their errors in formulas 1 and 2.

For any given survey estimate:

Similar to

Nonresponse Bias = Response -Nonresponse

Nonresponse Error = Nonresponse rate * Nonresponse bias (1)

Noncoverage Bias = Coverage -Noncoverage

Noncoverage Error = Noncoverage * Noncoverage bias (2)

etc..

Calculate

Non-Consent Bias = Responses of Consenters – Responses of Non-Consenters

Non-Consent Error = Non-Consent Rate * Non-Consent Bias (3)

These are all non-sampling errors that TSE focuses on and that can be relatively controlled by optimizing designs. The difference between respondents and non-respondents is called non-response bias. We expect that the difference between consenters and non-consenters data can be called non-consent bias (Formula 3). This paper focuses on the determinants of consent behavior, which helps researchers to see the differences

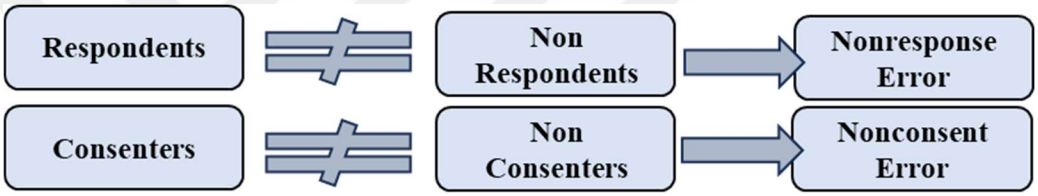
between consenters and non-consenters that lead to nonconsent errors. In general, it is easier to access information from consenters than from non-consenters because their data is not linked. Typically, not all sampled participants provide consent. There are consenters and non-consenters. Consenters' survey data is linked to administrative data. Failure to obtain a positive response to consent to record linkage can indirectly lead to bias in a survey once the survey data is linked to the administrative data. The first problem arises when non-consenters do not have the same characteristics as consenters, which is likely to lead to biased survey estimates (Kho et al., (2009); Sakshaug and Kreuter (2012); Sakshaug et al. (2017)). The second problem with nonconsent is that it reduces the effective sample size because fewer units are linked to the administrative data than expected, as confirmed by the survey. Overall, as the sampling variance increases, estimates become less accurate. If it is possible to predict the response rate in advance by taking advantage of a survey conducted in the past on a similar topic, the sample size of the new survey can be recalculated and increased accordingly. Increasing the sample size reduces the sampling variance, thereby correcting for randomly distributed nonconsent. However, if the non-consent is systematic, it does not reduce the bias. To sum up, from a TSE perspective, sampling error caused by non-systematic consent can be adjusted by increased sample size. For systematic consent, biases are to be measured as nonconsent bias, and necessary adjustments as nonresponse, post-survey, etc. adjustments are evaluated.

In this paper, we do not calculate or evaluate the nonconsent error or bias but try to understand the mechanisms that lead to nonconsent. It is aimed to review surveys asking for consent for data linkage in the literature internationally and to find out what has been influential for survey respondents to permit combining their survey data with other registered records, in other words, what are the significant factors of giving consent for linkage structured as survey design and consent characteristics variables in this research. Naturally, the likelihood of getting the same estimates for respondents and non-respondents, consenters and non-consenters is almost impossible. The intention is always

to keep the differences between them at a minimum. To do this, research is done on the factors that cause errors and ways to differentiate as much as possible under the TSE framework. The difference between estimates of respondents and nonrespondents gives nonresponse error, similarly, the difference between consenters and nonconsenters constructs gives the nonconsent error (Figure 2.2.).

Figure 2.2. Error Types Explained under Total Survey Error Perspective

Source: Application of Survey Errors Calculation Approach to Consent (Brand, 2024)



In addition to the leverage salience theory explained in the literature review, the TSE framework, which is structuring the development of other types of errors in a survey that asks for permission to consent data linkage to the survey respondents, are those theories that could be associated with the thesis arguments. Given those, the focus of this dissertation is to find out identifying factors to be a consenter or non-consenter. That way some guidelines on designing surveys for a minimum nonconsent error can be developed.

CHAPTER 3. SYSTEMATIC REVIEW OF CONSENT FOR DATA LINKAGE

This section of the thesis uses the methods of a systematic review to find resources that explain the application of obtaining consent to link data to administrative records. A systematic review is a broader definition adopted by the Cochrane Collaboration (1993) to include meta-analysis, i.e. the use of statistical techniques. However, meta-analysis is not necessarily used to analyze and summarize the results of systematic reviews. Meta-analysis is described in several ways (Borenstein et.al., 2009; Lipsey and Wilson, 2001; Littel et. al, 2008). One of these is a set of quantitative methods for reviewing literature, evaluating previous research & identifying gaps in knowledge found in published literature. Another description is a systematic combination of information from different sources. It is also referred to as a statistical summary of what is common and what needs to be analyzed differently. Or, as this research approach, a systematic review (or research synthesis) using statistical techniques is a review of a field of study based on a formulated research question that uses systematic and explicit methods to identify, select, and evaluate relevant research. To understand consent behavior, primary studies were systematically collected to identify surveys where consent for data linkage was obtained. Similar to the approach of Cornesse and Bosnjak (2018), a snowball search was also used by checking the literature section of key articles for relevant others. Once accessed to this literature, readings were made in detail to collect characteristics of surveys and their consent.

This dissertation examines the characteristics of surveys that may be influential in explaining consent behavior for linking survey and administrative data. Based on previous survey research, the potential impact of various survey design features and consent request methods on consent to link data is estimated. These assessments are based on data collected through systematic reviews of previous survey studies. After reviewing the literature, including questionnaires, publications, and methodological documentation, it became clear that certain survey features have a significant impact on obtaining consent. In order to analyze these characteristics, the characteristics outlined in Section 3.4 are used

to generate categorical variables that are tested through statistical models. For the investigation of consent for data linkage, the consent rate is used. The consent rate, a continuous dependent variable, is defined as the percentage of respondents who authorized the linkage of their survey responses to registered data over survey respondents during that survey period.

3.1. Data Sources

To construct the data for this thesis, surveys asking for consent to link data were examined. "The Preferred Reporting Items for Systematic Reviews (PRISMA)" checklist (Moher et.al. 2010) was used as a tool to identify these surveys. This checklist is a specific procedure with systematically identified steps that are useful for organizing survey research and converting observations into a dataset. The PRISMA checklist includes steps for identification, screening, eligibility, and enrollment. As a first step to identify studies, bibliographic databases such as Google Scholar, JSTOR, EBSCO, Medline, ProQuest, SAGE, ScienceDirect were searched using the keywords. Keywords appear partially. They have also been considered in combination with other words. The search is limited to full-text manuscripts in English with no publication date restrictions. Because the final search was conducted in November 2023, the cutoff date was set to include publications through October 2023. Unpublished studies and grey literature were reviewed if they were referenced by another study in the pool. In the second step, the screening found duplicates within these studies and removed them. The full-text articles were then evaluated for eligibility according to the inclusion and exclusion criteria. This determined the final number of articles to be included in this dissertation review.

Because these studies are publicly available and do not contain individual-level information, there is no ethical concern about violating any regulations.

In summary, the PRISMA steps to follow (Figure 3.1):

1. **Identification:** Studies identified through Google Scholar, other databases, and other sources such as snowballing, conferences, reports,
2. **Eligibility:** Studies assessed and excluded,
3. **Inclusion:** Studies included which held the criteria defined

3.1.1. Inclusion and Exclusion Criteria

A prior eligibility criterion for all survey data found and collected is that the researchers can calculate consent rates or have enough information to calculate consent rates for data linkage. Consent to link the data should be requested and the article should be written in English.

The data naturally included different types of surveys such as longitudinal surveys, cross-sectional surveys, and experimental research. The basic criterion for the data was to obtain information over some time about the respondents who were asked the consent question for the first time in that particular survey. Therefore, for survey types where respondents were followed over time, such as longitudinal surveys, the first wave of the survey was used. However, if there was a newly recruited refreshment sample or a methodological change in the survey that was likely to affect consent behavior over the years, the first wave of that new period was also included in the dataset. “The Panel Study of Income Dynamics (PSID)” and the “British Household Panel Study (BHPS)”, longitudinal surveys, time to time with refreshment samples are examples of such surveys.

Due to the sampling design of cross-sectional surveys, different groups of respondents may participate in each period. Therefore, consent for data linkage was required from these participants, and consent rates were recorded in each period. Also, experimental research, which is processed by using the survey data with a consent research question, and variables, are involved in the data. For example, in Sakshaug et al.'s (2017) “Legitimation of Inequality over the Life Span (LINOS)” study, the entire employment survey sample was randomly assigned to interviewer-administered (face-to-face) or self-

administered (mail/web) interviews that included consent questions linked to federal employment records. The purpose of this experimental study was to determine the differences in consent rates and consent bias among the survey modes tested for the full sample when requesting consent for record linkage. Therefore, they were treated as separate data observations in the data.

The exclusion criteria were as follows. If a study asked for both survey participation and linkage consent at the same time, it was excluded. In other words, if they were part of the same request, both the informed consent to participate in the survey and the consent to link the data at the same time, those studies were not included. As could be seen from the method itself, there was no way to calculate consent rates and evaluate consent behavior in such surveys. This is because the consent rate is the proportion of respondents who say "yes, I consent" compared to the total number of respondents. The consent question is only asked of those who have already agreed to participate in the survey, i.e., survey respondents. In other words, they are those who have given informed consent to survey participation earlier.

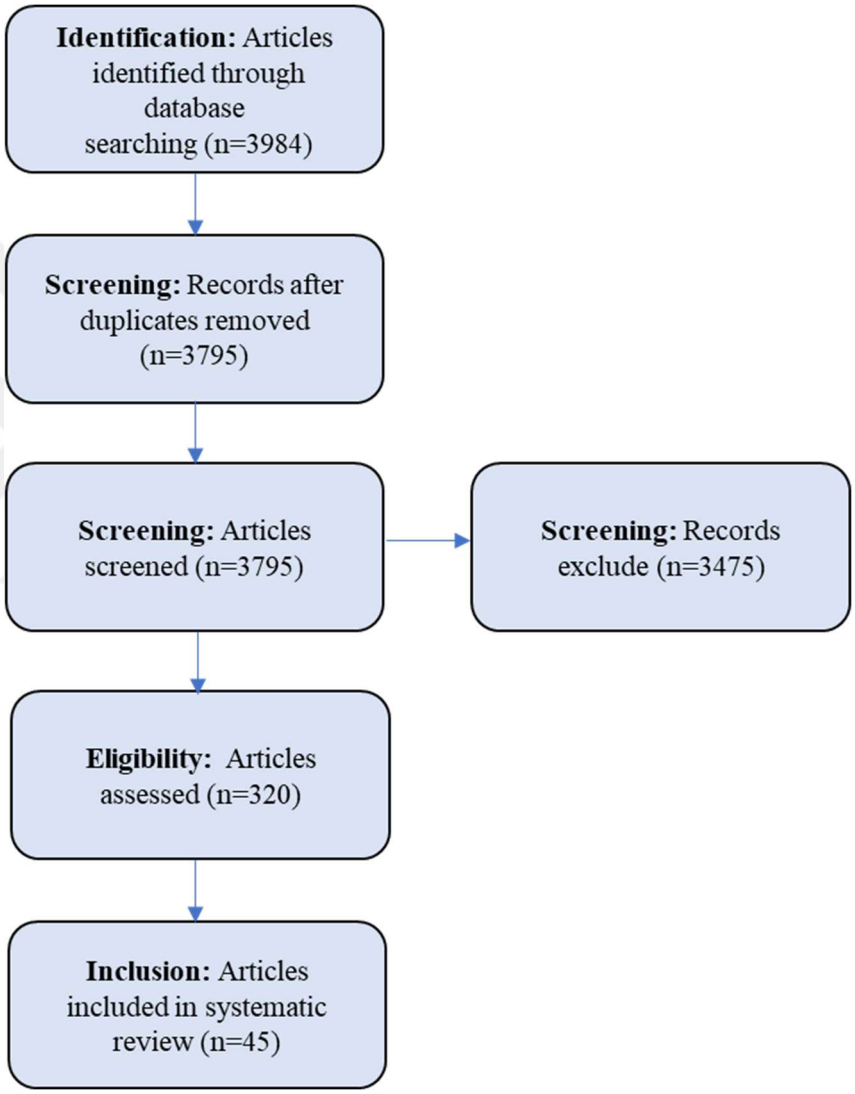
As expected, more than one article was examined using the same survey data. The strategy was to mention those articles with more information that were examined when creating variables. The others were used as a supplemental resource for missing information about the survey included. There have also been cases where the article was first studied as a working paper and later published in a journal. In this situation, published research was preferred only if sufficient information was available. After all these steps, 45 articles were included in the systematic review. Table 3.1. shows the surveys found in those studies for the given survey years and the country of origin. The author represents the paper published or not published, report by the indicated first author. For some authors in the same year, multiple publications are possible. The survey indicates the name of the study which is experimental, cross-sectional, or panel.

3.1.2. Data Preparation and Eligible Surveys

The electronic search produced an initial list of 3984 references. 189 out of 3984 articles were eliminated because they were exact duplicates of the same article in different article references. After reviewing the remaining 3795 articles, several situations were identified. Linkage error has been studied when there are different matching procedures implemented when reaching out to larger administrative data files with more variables needed. There were some other consent forms, that is informed consent, asking for permission to participate in the survey. Physical measurement permissions for health research and health surveys had a different type of consent. As a result, 3475 articles were excluded and 320 were in scope. Of these 320 articles, 275 cases were removed because they were duplicates or asked both informed consent and linked consent questions at the same time in the survey. Analysis of the titles and abstracts, and further examination of the reference lists in the retrieved papers and other sources, led to a preliminary list of 45 potentially relevant articles in "Germany, the United Kingdom, the United States, Canada, Australia, and New Zealand". Through the systematic review of articles, papers, and reports in the literature, the information of 50 surveys were collected from 45 articles.

Figure 3.1. Article Identification and Eligibility Assessment, PRISMA Flowchart

Source: Moher et.al., 2010



Using the data set construction strategy, these studies retained 45 articles and 128 survey data observations, in other words, these consent rates were extracted using the eligibility criteria described above. As expected, all of these surveys used different methodologies. The methodologies varied depending on the purpose of the survey research. Each of these studies found through the use of keywords was examined to determine whether it was a methodological report of a repeated cross-sectional or panel study, or a specific experimental project of a continuing or newly constructed survey project for this research. All studies reporting surveys with consent for linkage in all areas were included, including experimental research on these surveys. Although only English language resources were included in the dataset, German language resources were also used, when necessary, as a source of missing information needed to construct variables. Within the abstracts reviewed, if a study with consent to link was found, but not all information could be accessed, then other resources, reports, or as a last step, authors were contacted to obtain the relevant information. Three authors were contacted. The government agency was contacted for missing data for one survey.

3.1.3. Data Preparation: Surveys Researched

Keywords used in searches include "consent," "permission," "linkage," "authorization," "link," "administrative data," "registered data," and similar combinations of all. Table 3.1. below shows the published or non-published research that was accessed to prepare the data.

Table 3.1. Included Studies

Author	Survey	Survey Years	Country
Baghal 2014	"UK Household Longitudinal Study (UKHLS"	2009	“UK”
Carter 2010	"Survey of Families, Income and Employment (SOFIE)"	2004	“New Zealand”
Sakshaug 2012	"Labor Market and Social Security (PASS)"	2007	“Germany”
Woolf 2000	"Health Assessment Survey (HAS)"	1999	“US”
Young 2001	"Australian Longitudinal Study on Women’s Health (ALSWH)"	1997	“Australia”
Tate 2006	"Longitudinal Millenium Cohort Study (MCS)"	2001	“UK”
Hajizadeh 2021	"Canadian Community Health Survey (CCHS)"	2021	“Canada”
Klassen 2005	"Neonatal Intensive Care Unit Follow up Study (NICU)"	1996	“Canada”
Klassen 2005	"Healthy Children Study (HCS)"	1996	“Canada”
Harris 2005	"Elderly depressive symptoms survey (UKElderlySymp)"	2003	“UK”
Baker 2000	"The Living with Asthma measure (AsthmaQ)"	1998	“UK”
Baker 2000	"The Seattle Angina (SeattleQ)"	1998	“UK”
Soldo 1997	"Asset and Health Dynamics Among the Oldest Old (AHEAD)"	1993	“US”
Jay 1994	"Health Field Study (HFS)"	1993	“US”
Cleary 1984	"One time study (Cleary)"	1982	“US”
Fulton 2012	"Survey of Income and Program Participation (SIPP)"	2008	“US”
Sakshaug 2012	"Health and Retirement Study (HRS)"	1994	“US”
Fulton 2012	"Second Longitudinal Study of Aging (LSOA II)"	1994	“US”
Fulton 2012	"Medical Expenditure Panel Survey (MEPS)-Health Care (HC)"	1996	“US”
Fulton 2012	"NHANESIII"	1991	“US”
Fulton 2012	"NHANES I Epidemiologic Follow up Study (NHEFS)"	1982	“US”
MacDonald 2023	"Avon Longitudinal Study of Parents and Children (ALSPAC)"	2021	“UK”
Taylor 2007	"English Longitudinal Study of Ageing (ELSA)"	2002	“UK”
Fulton 2012	"National Health Interview Survey (NHIS)"	1997-2018	“US”
Fulton 2012	"National Health and Nutrition Examination Survey (NHANES)"	1999 2001 2003 2005 2007 2009 2011 2013	“US”
Fulton 2012	"National Immunization Survey (NIS)"	1997-2017	“US”

Table 3.1. Included Studies (continued)

Author	Survey	Survey Years	Country
Fulton 2012	"Panel Study of Income Dynamics (PSID)"	2005 2007 2009 2011 2013 2015 2017	"US"
Fulton 2012	"Residential Energy Consumption Survey (RECS)"	1987 1990 1993 1997 2001 2005 2009 2015	"US"
Dahlhamer 2007	"National Health Interview Survey (NHIS)"	2007	"US"
Jenkins 2004	"A follow-up survey ISMIE"	2001	"UK"
Korbmacher 2013	"Survey of Health, Ageing and Retirement in Europe (SHARE)"	2008	"Germany"
Sakshaug 2016	"Four-wave panel survey of employees (2007-2011) (WeLL)"	1993 2002 2007	"Germany"
Sakshaug 2012	"Health and Retirement Study (HRS)"	2008	"US"
Sakshaug 2013	"Wording Experiment (Exp1)"	2011	"Germany"
Sakshaug 2013	"Placement Experiment (Exp2)"	2011	"Germany"
Sakshaug 2018	"Job Vacancy Survey (JobVacSurv)"	2014	"Germany"
Sakshaug 2017	"Legitimation of Inequality over the Life Span (LINOS)"	2012	"Germany"
Sala 2010	"British Household Panel Study (BHPS)"	2008	"UK"
Thornby 2018	"Next Steps Age 25 Survey (NextSteps25Surv)"	2015	"UK"
Smith 2021	"General Social Survey (GSS)"	2018 2021 2022	"US"
Antoni 2011	"ALWA Survey (ALWA)"	2007	"Germany"
Eisnecker 2017	"IAB-SOEP Migration Sample (IAB-SOEP)"	2013	"Germany"
Mostafa 2016	"Longitudinal Millenium Cohort Study (MCS)"	2015	"UK"
Sakshaug 2014	"Web Survey Random Experiment (WebExp)"	2012	"Germany"
Kreuter 2016	"Maryland Residents Random Experiment (MarylResExp)"	2001	"US"
Knies 2014	"1958 National Child Development Study (NCDS)"	2008	"UK"
Kuh 2011	"National Survey of Health and Development (NSHD)"	1946 1969	"UK"
Partin 2008	"Randomized control trial on a postal health survey (RTP)"	2001	"US"
Pascale 2011	"Survey of Health Insurance and Program Participation (SHIPP)"	2010	"US"
Sakshaug 2019	"Experiment using IAB Panel (ExpIAB)"	2013	"Germany"

3.1.4. Examples of Surveys with Consent to Data Linkage Question

Using the characteristics of each survey as input, the relevant variables are constructed and those are explained in Section 3.2. In this section, there are two examples of surveys with consent for data linkage questions in the data. They are "Health and Retirement Study (HRS)" and "Job Vacancy Survey (PASS)".

3.1.4.1. Example "Health and Retirement Survey (HRS)"

The HRS is a longitudinal panel study conducted by the University of Michigan since 1992 and supported by the National Institute of Aging and the Social Security Administration. The HRS includes several physical measures and biomarkers, as well as a self-administered psychosocial questionnaire. The consent question at the end of the survey was explicitly asked, with a return of consent rate of 67.8% in 2008.

Consent question wording: *"We would like to obtain a history of your earnings and any benefits from programs administered by the Social Security Administration applied for or received through 2023. Since most people cannot recall this information very well, we are asking for your permission to obtain from government records the following: 1) Your earnings reported to Social Security. 2) Any information about benefits from programs administered by the Social Security Administration applied for or received through 2023 (HRS, https://hrsdata.isr.umich.edu/data-products/2008-hrs-core)".*

Sakshaug et al., (2012), in their research, explained their experimental research for 2008 HRS with several cases of hypothesis whether respondents are willing to consent to earnings and benefits data linkage requests. They looked at a variety of characteristics, including privacy, survey resistance, financial unknowns, and in-person participation, to find out how effective they were on respondents' decisions about consenting to data linkage.

3.1.4.2. Example "Labor Market and Social Security (PASS)"

PASS uses two modes both CATI and CAPI. It is an annual German study conducted by the "German Institute for Employment Research" (Trappmann et al., 2010). It is a household panel survey focusing on welfare, labor market, and poverty issues in Germany. The first wave was implemented from December 2006 to January 2007. The survey response rate was 26.7%, while the consent rate was high at 78.4%.

Consent question wording: "*To keep the interview as brief as possible, the Institute for Employment Research in Nuremberg could merge the study results with data about your times of employment, unemployment or participation in measures by the employment office (Arbeitsamt). For the results of this study, it would be a great advantage. For reasons of data protection this cannot be done without your agreement, which I kindly ask you to provide. This is of course just as voluntary as the interview you are so kind as to give us. Of course, you may withdraw your consent at any time. It goes without saying that all rules of data protection and of the de-personalization of the results reported apply to these additional data as well. So may I write down your answer: Do you agree to the use of this additional data?" (PASS, https://fdz.iab.de/en/pd_hd/panel-study-labour-market-and-social-security-pass-version-0618-v1/)*

3.2. Data Preparation: Constructed Variables

This section discusses the variables that were created to be examined in statistical models to explore their effects on explaining consent rates, and where the data to create these variables came from. The variables are survey and consent design characteristics in categories which are also summarized in Table 3.2. The variables used in this study come from categorized discrete characteristics of survey and consent features, except survey response rate, which is a continuous variable. The reasoning behind the creation of these categories is discussed in this section. Table 3.2. shows the categories of variables created.

Survey Consent Rate. The consent rate, a continuous variable, is the percentage of respondents who permit their survey responses to be linked to their registered data out

of the total survey respondents. In other words, the consent rate is a type of effect size in meta-analysis terminology, a dependent variable of this study that can be measured across different survey studies. There is a section in these surveys where the linkage consent question is asked via a written or verbal consent form. These forms explain in detail to respondents the intention to link survey data with administrative data and ask for their responses. Their response is expected to be an approval or disapproval of the linkage consent. In addition, their concerns or questions are addressed to ensure that the purposes of the linkage are satisfactorily understood.

Survey Topic. For each type of consent request, the consent rate for different survey topics shows variability across surveys. In most cases, health surveys have a higher rate of informed consent to participate in the survey (Voogt and Van Kempen, 2002), but according to an experiment conducted by Keusch et.al (2019), the topic does not affect the associated consent. In this study, surveys with topics such as education, income, health, and others were categorized as a variable.

Learning the survey topic is like a first impression for respondents, where they begin to form some ideas or expectations about what comes next in this survey communication. Therefore, if the topic is somehow appealing or relevant to them, it can be an asset to their survey participation behavior and even consent for data linkage for some reason (Groves, Presser, and Dipko 2004). There is also evidence (HMRC, 2010; Snijkers et al., 2013; Snijkers, 2018) that the topic of the survey is important. In the case of multiple topics, such as in the British Household Panel Study, New Zealand Survey of Families Income and Employment, the strategy is to take the highly weighted topic, evaluated by the number of relevant questions, as the survey topic for this research.

Survey Country. Given the number of factors that are likely to be at play in countries, it would be very ambitious to claim that one country is more or less likely to consent to data linkage than another. Through their socio-economic structures, cultural characteristics, and thus their laws, they develop different attitudes towards consent, privacy, and data sharing issues. Therefore, it is difficult to assess whether the variability

in the laws and social and cultural conditions related to consent causes a major difference in the consent rates of "Germany, the United States, Canada, Australia, New Zealand, and the United Kingdom" in this study.

Survey Year. Compared to previous years, the linking of survey and administrative data has become more common, and more sophisticated procedures have been developed to obtain consent. Although respondents have been more cooperative, it is unlikely that increased awareness of privacy and confidentiality issues will lead respondents to be more secure with their data, resulting in lower consent rates for data linkage. It is not easy to know whether surveys from older or later years have higher consent rates.

Survey Type. In cross-sectional surveys, consent for data linkage is asked at each period of the survey; in panel surveys, it is asked to the sample constructed in the first wave of the survey respondents to consent. The main difference between these survey types is that cross-sectional surveys examine the situation at one point in time, whereas panel, longitudinal, or cohort surveys measure changes in the same sample over time to make inferences about a population of interest. As clarified in the consent question, they may change their decision at any time in future waves of the survey once they are reminded of their consent to data linkage.

Panel respondents are expected to start with trust in the organization conducting the survey and build more trust over time (Sakshaug et al., 2012). However, respondents typically do not know or pay attention to whether the survey is panel or cross-sectional until the second wave. That is, it is unclear when and to what extent they are informed about the survey type, which may determine their decision to consent to data linkage. This raises the question of whether or not survey types differ in their data linkage consent rates.

Survey Target Population Age. There is inconsistency across studies in how some survey characteristics affect consent behavior. Participant age is one effective characteristic, possibly in both positive and negative directions. For example, some studies find that older respondents are more likely to consent to data linkage (Bryant et al., 2006; Dunn et al., 2004), while others find that younger respondents are more likely to consent

to data linkage. (Hung et al., 2007; Yawn et al., 1998). In addition, some studies have found that age is not a significant factor in determining consent for data linkage (Buckley et al., 2007; Harris et al., 2005).

Survey Mode. The need to explain data linkage to survey respondents by providing as much information as possible with clearer explanations is emphasized (Thornby et al., 2017), which is guaranteed in interviewer-administered surveys. There is an opportunity for participants in an interviewer-administered survey to ask clarifying questions; having an interviewer is a highly preferred option for these groups of respondents, as observed in an Understanding Society longitudinal survey. (Jäckle et.al, 2018). In a randomized mode study of the Legitimation of Inequality over the Life Span (LINOS) panel survey in Germany, the CAPI survey mode had a linkage consent rate about 40 percentage points higher than the self-administered survey mode (Sakshaug et al., 2017).

In this study, to determine the effect of personal contact with interviewers on consent behavior, surveys were categorized as CAPI, CATI, self-administered, dual mode (CATI and CAPI or CAPI and Web), and sequential mode. Sequential mode is when a mode type, such as self-administered, is first implied and then telephone mode is sequentially implied to the same sample after the previous mode failed to reach the respondent.

Survey Sponsor. The survey sponsor is usually identified to respondents during the initial contact when the purpose of the survey is explained and the consent forms request a response to the survey. In general, government-sponsored surveys may be perceived as compulsory to some degree, probably depending on the governmental system in different countries. On the other hand, non-government surveys tend to give respondents more flexibility in making their own decisions about whether or not to participate. Assuming that the government-sponsored survey happens to have a positive return on response (Heberlein et al., 1978), some research shows that government-sponsored surveys achieve higher response rates. (Linsky, 1975; Goyder, 1985; De Leeuw

& Heer, 2002). It is believed that respondents may have stronger bonds of trust and respect for their governments, or feel a sense of obligation to participate, as well as to consent to data linkage. When asked to consent to linkage, respondents may not want to share their private information with the government unless there is a compelling situation, similar to the response behavior observed in the Groves et. al. study (2008). On the other hand, the government, which is the beneficiary of the survey project, does not necessarily have direct access to the private data, since the focus of policymakers is the output that is analyzed using these data.

Survey Response Rate. Participating in a survey and allowing the linking of one's registered information are not necessarily similar issues when both are investigating their questions. Participating in a survey is much easier to accept than sharing relatively private registered information. Respondents may view participating in a survey as beneficial to scientific research to the extent that their confidential information is not shared through linkage. Also, some respondents may feel comfortable doing both. That is, consent to take a survey and consent to link data may not be relevant at all. Sometimes, respondents with lower response rates are likely to be more selective and cooperative. Because these respondents are a group that fully trusts and engages with the survey, a higher percentage of them will naturally agree to link. On the other hand, research (Groves, et al., 2004) discusses that the contribution of factors such as effective interviewers, advance letters, and refusal conversion efforts to unit response and willingness to consent to administrative data linkage are predicted to be positively related.

Survey Consent Request Type. There are two approaches, either explicit (active) or implicit (passive), to obtaining survey consent from participants. In the explicit consent approach, subjects are asked to sign and return a consent form documenting their permission for their contact information to be transferred from administrative sources to third-party data collection agencies. In the implicit consent method, however, permission is granted if no action is taken. In other words, the signed opt-out form is not returned; the

respondent is responsible for opting out at some point in the survey process. Data linkage consent is the default option.

In general, explicit consent is a better choice to indicate the respondent's direct thoughts about participating in a survey, although it may lower consent rates and increase bias in survey estimates. Sakshaug's (Sakshaug et al., 2016) research found that explicit consent rates are lower than implicit consent rates. However, for consent, non-response, and survey estimation bias, the situation is reversed, with implicit consent procedures yielding less bias than explicit procedures. Das and Couper (2014) test implicit consent for linkage in a probability-based Internet panel. After observing the results of an implicit consent text through qualitative research they conduct an experiment to test the effects of request type and content. Presenting an implicit consent statement does not make respondents more hesitant about privacy, and confidentiality.

Linkage Request Placement. There is no definitive agreement on where to place the consent question in the survey, although most studies have placed the consent question closer to the end of the questionnaire. Linkage consent is perceived as a sensitive issue, and interviewers must first establish rapport and trust with respondents before asking for consent to link data. It is better to ask this question closer to the beginning of the survey (Sakshaug et al., 2013).

Linkage Request Records. In most surveys, the records requested for consent linkage are on the same subject as the survey subject. Although the need for linkage to administrative records arises mainly in employment and health surveys, there are other records, such as education, income, and tax credit records, that may be requested for linkage in the context of surveys on the same or different topics. In the first wave of the UK Household Longitudinal Study (Baghal et al., 2014), multiple data linkage was used to identify differences based on the type of records requested and consistency across different consent questions. Household composition, interviewer information, respondent characteristics, and survey environment are other components of the model to observe differences between health and education records. It has been found that people are less

willing to consent to the linking of health records than to the linking of education records. However, the elderly population, who need more help from the medical system, may be more likely to respond positively to a consent question. The study by Sakshaug, et.al. (2012) suggests that older respondents may be more cooperative, resulting in higher consent rates with fewer privacy concerns than younger adults in the Health Retirement Survey. However, it is questionable whether the consent rate would be higher if the records were about health or not.

Consent Wording. The wording of the consent question and the information provided to explain the consent to data linkage can play a significant role in respondents' attitudes. (Sakshaug et al., 2013). It is well known that a strong guarantee of data confidentiality leads to higher response rates. However, this effect can also be reversed (Singer et al., 1999). In general, by emphasizing the benefits of linkage to respondents, research means to help meet a similar experiment in a Web-only survey yielded higher consent rates with time-saving argument-indicated wording than neutral framing wording condition (Sakshaug and Kreuter 2014). In addition, Kahneman and Tversky (1979, 1984) demonstrated in their research that people are risk averse. Under loss framing, respondents learn about disadvantageous situations in the case of nonconsent and are more likely to consent. As research shows, consent framing varies. Therefore, it is discussed that framing the consent question neutrally versus negatively or positively will have different influences on respondents to give a positive return.

Survey Consent Signature and Identity Information. The decision to sign a linkage consent form can also be examined through theories of informed consent for survey participation, with the same concerns about privacy and confidentiality. If the survey is on a sensitive topic that includes questions about income, finances, or blood tests, the situation may become even more difficult as they are informed that their data can be much more easily linked to administrative data. ("Jenkins et al. 2006; Sala et al. 2010"). If respondents behave in the same way as they do when consenting to participate in a survey with a signature, then requiring a signature for linkage consent may be

disadvantageous compared to not requiring a signature. In addition, asking for identity information may make it more difficult to obtain a positive consent response. Related to this, either signature or ID; at least one; both categorized as variable.

Table 3.2. Constructed Variables Definitions and Categories

Variables	Definition	Categories
Survey Topic	The main subject of the survey. If more than one subject, which is relevant to the consent question is selected.	1: Other 2: Economic 3: Health
Survey Country	The country, which the survey is implemented	1: Australia, NZ, UK 2:US, Canada 3: Germany
Survey Year	The time when the survey is on the field.	1: Year>=2010 2:2001<Year>2009 3: Year <=2000
Survey Type	the frequency of the survey on the same respondents selected.	1: Cross-sectional 2: Panel, Longitudinal, Cohort
Survey Mode	Data collection method	1: Self-administered 2: CATI only 3: Other (Dual, Sequential) 4: CAPI only
Survey Target Population Age	Targeted age group in the sample of the survey.	1: All Age Groups 2: Age<18 3:18<=Age
Survey Sponsor	The organization where the costs of the survey are sponsored by.	1: Nongovernment (Academic, Private) 2: Government
Survey Consent Request Type	How the consent request is exposed.	1: Implicit/Passive/ Optout 2:Explicit/Active/Optin
Linkage Request Placement	The location of data linkage requests questions in the survey.	1:at the End 2:at the Middle &Before Consent
Linkage Request Records	The subject of administrative records asked for data linkage.	1: Other 2: Economic 3: Health
Consent Wording	Whether consent question wording is positive/negative framed or neutral	1: Framing 2: Neutral
Consent Signature and ID information	After consent approval by the respondent, which one of signature, identity information, or none are required.	1: Signature and Idinfo 2: Signature or Idinfo 3: None required
Survey Response Rate	The proportion of respondents over the sample population	Continuous

3.3. Statistical Methods

The methodological approach of the macro part of the dissertation is a systematic review to create a data set with the indicated guidelines and analyze this data. As explained in the previous Chapter 2, the optimal design of the data is highly emphasized with detailed screening, eligibility, and inclusion and exclusion steps which are implemented using PRISMA guidelines. For this research, the consent rate, dependent variable, i.e. the ratio of respondents to the total number of respondents in the survey, is modeled with the best possible approach. In building the models, the survey design and consent variables are described as independent, exploratory variables and analyzed. For each independent variable, categories are created based on the number of observations in each category and those that show differences in consent rates based on the literature review. The reference category is chosen as one that is frequently compared in the literature for its response to consent for data linkage (Table 3.3.). Initially, descriptive statistics including frequencies, means, medians, and correlations were evaluated for the survey response rate and the consent rate for all variables. The frequencies and mean consent rates of each category are discussed for the survey and consent request variables by each design feature of interest. Naturally, there is variation between studies, but the systematic review allows for heterogeneity by assuming that the underlying effects follow a normal distribution. Because systematic reviews systematically combine data from several selected studies to develop a single conclusion, they are expected to have greater statistical power. There is an increased number of subjects, greater diversity among subjects, or accumulated effects. As an alternative Kruskal Wallis test is used to compare each variable category's survey consent rate averages. Kruskal Wallis test is a nonparametric approach to a one-way analysis of variance (ANOVA) test.

The consent rate is then modeled as a continuous dependent variable using multiple regression analysis with discrete variables plus the continuous response rate. Another model stepwise regression analysis is implemented and it is found that the same variables with similar significances explain the consent rate.

For all statistical analyses, R. 4.2.2 for Windows software for statistical computing is used.

3.3.1. Survey Consent Rate Calculation

The survey consent rate, the dependent variable, is calculated as the number of respondents who consent to data linkage as a percentage of the total number of survey participants in the given survey period.

CR: Consent Rate

$$CR = \frac{number\ of\ respondents\ consent\ to\ data\ linkage}{total\ number\ of\ survey\ respondents} * 100 \qquad (4)$$

3.3.2. Survey Response Rate Calculation

The survey response rate, the only other continuous variable after the consent rate, is calculated as the number of respondents who agreed to answer the survey as a percentage of the total number of sampled respondents selected from the target population. For this thesis purposes, due to the inclusion and exclusion criteria explained, the only surveys eligible are those for which consent to participate is not obtained at the same time as the consent request for data linkage.

RR: Response Rate

$$RR = \frac{number\ of\ respondents\ accepted\ survey\ participation}{total\ number\ of\ sampled\ respondents} * 100 \qquad (5)$$

3.3.3. Kruskal Wallis Test

Kruskal Wallis is a nonparametric test that assesses whether the mean rank scores of a given variable are different between groups, in other words, nonparametric equivalent of one-way ANOVA. For this thesis, the question is whether there are significant differences in average survey consent rates between the categories of variables. The null hypothesis (H_0) is that the median survey consent rate is equal across variable categories. The alternative hypothesis(H_1) is that at least one of the mentioned variable categories' survey consent rate differs from the median of the other category.

The hypothesis is assessed by using test statistics, H as below.

$$H = \left[\frac{12}{N(N+1)} \sum_{j=1}^{c} \frac{T_j^2}{N_j}\right] - 3(N+1) \qquad (6)$$

Where j: 1,2,..4

N: total number of observations in each category

T_j: rank total for each category

N_j: number of observations in each category

3.3.4. Multicollinearity Check of Model Variables using VIF

Multicollinearity is assessed by calculating a variance inflation factor (VIF), which measures how much the variance of a regression coefficient is increased due to multicollinearity in the model.

VIF is a robustness test to find out if there is any multicollinearity between predictors of a fitted model. VIF calculates the extent of correlation between one predictor and the other predictors in a model. It indicates the increase in the variance of a regression coefficient in the model due to an existing collinearity. If it exists, it decreases the statistical significance of the independent variables.

If in a statistical model, two or more predictors are linearly related, collinearity in other words multicollinearity exists. Depending on the degree and reasons, statistical methods are sensitive to that. Due to two predictor variables collinearity, their effects cannot be distinguished which causes erroneous calculations. There is no zero collinearity between predictor variables. The collinearity of predictors is accepted to some degree. High collinearity leads to missing importance of predictions and imprecise estimates of β.

VIF is formulated below.

$$\boldsymbol{VIF_i} = \frac{\mathbf{1}}{(\mathbf{1}-\boldsymbol{R_i}^{\mathbf{2}})} \qquad (7)$$

where i = 1,2, 3,…. 13 independent variables

R_i^2 = unadjusted coefficient of determination to regress the i^{th} independent variable on the other independent variables

That says if R_i^2 is close to one then VIF will be large.

If VIF = 1 then the independent variables are not correlated

It is very common for VIFs greater than 1. That is there are at least two variables in multiple regression that are not orthogonal. That way, for each of these variables slope, SE, p-value alter due to shared variance between predictors.

As a rule of thumb, if the VIF calculated is 5 or below, it is not a concern. A VIF greater than 10 indicates high multicollinearity and starts to be problematic for the validity of results (Dormann et al., 2013; Hair et al., 2014).

3.3.5. Linear Regression Analysis

With the guidance of descriptives, some assumptions of linear regression models are checked. The first assumption is the existing linear relationship between the variables. Second the values of the dependent variable, y, and consent rate are independent of each other. Third, for any given value of X, the y values and the error terms are normally distributed. The last assumption is the residual variation is constant, i.e. they have an equal variance.

The general form for the model of the consent rate response in terms of predictors, and constructed variables is:

$$\boldsymbol{Y = f(X_1, X_2, X_3 , \dots\dots X_{13}) + \epsilon} \qquad (8)$$

f: *unknown function of predictors /independent variables*

$\boldsymbol{\epsilon}$ *: error in additive form*

$$\boldsymbol{Y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{13} X_{13} + \epsilon} \qquad (9)$$

where β_i *, i=0,1, 2, 3, ... 13 are unknown parameters.* β_0 *is the intercept term.*

As seen, in a linear regression model, parameters enter linearly but not necessarily the predictors.

As a result of the linear regression analysis, confidence intervals are commonly calculated to learn the values set to the estimates if repeated several times. For that, errors are assumed to be normally distributed and the number of observations is large enough, that the estimator is approximately normally distributed (the Central Limit Theorem applies). As a next step, a stepwise approach is applied to diagnose the results of linear regression analysis. Thus, there are two final models in which the results are comparable.

In Chapter 5, the results of this macro perspective on the consent behavior analyses are discussed in detail. Next, Chapter 4 explains the second part of the study with GSS data, as a micro perspective, and discusses our approach.

CHAPTER 4. MICRO APPROACH: INDIVIDUAL CONSENT BEHAVIOR IN GENERAL SOCIAL SURVEY

In the first part of this thesis, consent behavior from a macro perspective with 50 surveys on a variety of topics to learn about some determinants of consent approval for data linkage is investigated. In the second part, the aim is to look into the same subject more in-depth by using a specific survey as the "General Social Survey" data of the United States. This survey has been conducted by the "National Opinion Center at the University of Chicago" since 1972. GSS is run annually and data is publicly available. It does not have any real data identification information, which secures the privacy and confidentiality of those respondents. Therefore, it does not require extra procedures to ask for access.

After 1994, there have been innovations to the GSS. The core set of questions was reduced to permit the new type of mini-module style questions to start. That way important topics proposed by the research community and some experimental designs can be tried. Second, a split-sample design was started, consisting of two parallel subsamples with identical cores and different modules. In 1984, institutions from Australia, Germany, and Great Britain and the GSS agreed together to establish the International Social Survey Programme (ISSP). There were modules of ISSP experimented in the new design. GSS questions change in every round to reflect emerging issues such as the COVID-19 pandemic, crime, and political polarization.

In our investigation in the first chapter where a pool of surveys selected by systematic review was researched, analyzing respondent-level information was not possible. Differently in this chapter, with the selected demographic variables, the characteristics of respondents in the GSS, consenters, and non-consenters, were able to be analyzed for the year 2021. Trust with organizations as witnessed as one of the main factors in the first part will be looked into with more depth in this chapter.

4.1. GSS Data

Since 1972, the General Social Survey (GSS) has been a representative survey of U.S. adults that collects data on contemporary American society to monitor and explain trends in opinions, attitudes, and behaviors. Because the number of questions has remained the same for about 80 years, researchers can make comparisons between years. It is recognized as one of the most important sources of attitudinal and behavioral data explaining trends in the United States. Topics covered in this study include psychological well-being, social mobility, stress, and traumatic events. The main areas of the survey include socioeconomic status, social mobility, social control, family, race relations, gender relations, civil liberties, and morality.

Beginning with the 2018 survey, GSS decided to take advantage of data linkage and began asking consent questions for data linkage. During the pandemic, GSS was out of the field, so it missed the 2019 and 2020 years. It resumed in 2021. Since then, consent for data linkage has been asked each year in addition to the regular GSS questionnaire.

In the macro section of this dissertation, some indicators showed that trust in organizations, interviewers, and others can play a critical role in the decision-making process of respondents. The main reason for choosing the GSS for the micro part of this thesis is that it has a trust section with the sufficient number of questions. Secondly, consent for data linkage has started to be asked lately, a relatively new addition to the survey which will be more helpful in understanding current consent cases.

4.1.1. Survey Design of GSS

The sample consists of adults 18 years of age or older in the United States living in non-institutionalized housing at the time of the interview. Respondents within the household are selected using the last-birthday method. The field period is December 1, 2020, through May 3, 2021. Initially, an administrative mail will be sent to inform and collect data via the Internet. Also in this mail, they will be informed with a web extension to invite them to participate on the web. Both mail and web modes will be supplemented

by telephone. In 2021, the final sample size is 4032 completed within 27591 sampled. Thus, the response rate according to AAPOR measures is 17.4%. There are non-contingent prepaid incentives and contingent postpaid incentives. The questionnaire is available in English and Spanish.

The sample of GSS 2021 was released in three different batches, periods with different start dates between December 1, 2020, to May 3, 2021. 5,891 addresses from the NORC 2010 NORC National Sampling Frame and 4200 unclustered addresses from the United States Postal Service (USPS) Computerized Delivery Sequence File (CDS), in total 10091 addresses constitute the first batch. Batch two and three are composed of 10000 and 7500 un-clustered addresses from the CDS. A stratified un-clustered address sample was used for data collection. All batches supported the web instrument with a phone backup.

NORC selects the National Sampling Frame (NFAs) on a Decennial Census basis. USPS and Master Shipping Accounts (MSAs) form the first stage National Sampling Frame. 2010 NORC National Sampling Frame has 126 NFAs and the GSS sample uses 76 of those NFAs. Fifty-nine of those represent 1 percent of the US population with non-certainty and 17 of them are certainty, 41 percent of the US population. In the second stage, census tracts in urban areas and block groups smaller than census tracts in rural areas are selected. That requires a physical listing. The first and second-stage units are selected with probabilities proportional to size (a measure of size is decided by Census housing unit totals). In the third stage housing units are selected with an equal probability of selection method (EPSEM), that is there is the same probability of selection for each household.

The survey uses a ballot and form design to conduct question experiments and ask for more content of a smaller sample. Typically, there are 3 ballots (A, B, C) and 2 forms (X and Y). Content will usually be:

(1) on all ballots and forms (i.e., AX, AY, BX, BY, CX, CY),

(2) on all ballots but only one form (i.e., AX, BX, CX),

(3) on two ballots and all forms (i.e., AX, AY, CX, CY),

(4) on both forms of one ballot and one form of one ballot (i.e., BX, CX, CY).

If a respondent is not assigned to that form and ballot, then they have an inapplicable (IAP) missing value for those variables. For example, the variable "natspac (space exploration program)" is available on AX, BX, and CX (situation 2 above) while the variable "natspacy (space exploration)" (an experimental version of "natspac") is available on AY, BY, and CY. "natspac" is IAP for AY, BY, and CY and "natspacy" is IAP for AX, BX, and CX. Another example is a variable like "sexornt (Which of the following best describes you?)", asked as part of the High-risk Behaviors module, which in recent years has only been on one form of B (BY) and both forms of C (CX, CY) (scenario 4 above).

There are sections including questions on voting and political preferences, religious attitudes and behavior, class identification and economic well-being, subjective well-being, and social life orientations. However, due to the ballot design, the sample size for some of the questions decreases. Thereby there is a limitation on using some of those variables for analyses due to diminished sample size and comparability issues.

4.1.2. Question of Consent for Data Linkage in GSS

Consent Question:

"May we try to link government records with your survey answers?"

In Table 4.1. response options to the consent question of the GSS are listed.

Table 4.1. Response Options of the GSS Consent Question

0	Respondent consents to possible data linkage
1	Respondent does not consent to possible data linkage
NA (d)	Don't know (dk)
NA (i)	Inapplicable (iap)
NA (j)	I donot have a job
NA (m)	dk, na
NA (n)	No answer
NA (p)	Not imputable
NA (r)	Refused
NA (s)	Skipped on web
NA (u)	Uncodeable
NA (x)	Not available in this release
NA (y)	Not available in this year

The reserve code iap is the most frequently seen missing code in most of the cases. Respondents sometimes do not see the question due to either structural or personal factors. Structural factors are, for instance, if they are in ballot B then they will be marked inapplicable in ballot A and C. Respondents who are not married will be inapplicable for marriage-related questions and thus are assigned as inapplicable. The consent request question “admin consent” has 549 responses of inapplicable. Because it is one of the final questions asked. These 549 cases are complete. Completes in this survey are respondents who did not get to the end of the survey, but completed at least two-thirds of the full survey and did not receive the final questions.

GSS explores avoiding as much as possible the missingness in data by putting extra effort into information sheets. It informs its respondents on the details and purposes of data linkage by declaring an information page and steps one by one. Thus, data linkage is

mostly clear on respondents' minds. This information sheet includes responses to their relevant frequently asked questions (Appendix A) such as below:

"Why are you asking to link my data? Why is it important?
What kind of information will you be linking my data with?
How will my data be protected? How do I know my information will be kept safe?
How will my data be used? Who will see my answers?
Can I change my mind later about providing my permission?"

Asking for permission from respondents for data linkage has been a relatively new application for GSS since 2018. The table explains the consent rate percentages for the years implied. N is the number of sampled survey respondents in that period.

Table 4.2. Percentage of Responses about Consent for Data Linkage

Year	Yes	No	N
2018	53.4%	46.6%	2339
2021	44%	56%	3483
2022	48.5%	51.5%	3422

4.2. Derived Survey Variables

GSS has many questions that help to look into certain characteristics identified at the macro level of this dissertation. Therefore, questions such as voting, political preferences, religious attitudes and behavior, economic subjective well-being, and social life orientations are divided into three groups, and analyses are run accordingly.

The first set of variables indicated in Table 4.3. represent the relevant question of demographic background with category frequencies. These variables with these categories are created by using the responses to the relevant questions. To our interest in observing differences between certain categories and the size of categories, variable levels are

merged and regrouped to name these study variables. While forming those categories, frequencies of levels are taken into consideration in specifying the cut-off values. All the variables are discrete except age being continuous.

4.2.1. Demographic Background Questions

Below is the wording of questions indicated to be used for constructing the relevant variables in Table 4.3. Frequencies of each category are documented in Table 4.3.

Age “How old are you?”

Sex “What is your gender?”

Marital: “Are you currently married, widowed, divorced, separated, or have you never been married?”

Race: “What race do you consider yourself? *Record verbatim”*

Birthcountry: “Were you born in this country?”

Degree: “*Record* school degree”

Income: “In which of these groups did your total family income, from all sources, fall last year before taxes, that is?”

Workself: “Are/Were you self-employed or do/did you work for someone else?”

Region: “*Record* region of interview”

Health: Would you say your own health, in general, is excellent, good, fair, or poor?

Children: “How many children have you ever had?”

Table 4.3. Demographic Background Variables

Variable	Level	Category (ref:1)	Frequency Percent (%)
Age		“Above 18 years old”	Continous
Sex	1	"male"	44.1
	2	"female"	55.9
Marital	1	"married"	49.7
	2	"Not married"	26.1
	3	"Never married"	24.2
Race	1	"white"	78.2
	2	"black"	11.6
	3	"other"	10.2
Birthcountry	1	"yes"	88.8
	2	"no"	11.2
Degree	1	"High school and below"	46
	2	"Above high school"	54
Income	1	"Under 25K"	20
	2	"25K and above"	80
Workself	1	"self-employed"	11
	2	"Employed by someone"	89
Interviewregion	1	"West Coast"	23.1
	2	"Midwest"	24.6
	3	"East Coast"	52.3
Health	1	"excellent"	20.8
	2	"good"	56.3
	3	"fair"	19.2
	4	"poor"	3.8
Children	0	"none"	29.2
	1	"One or two children"	45.1
	2	"three or more children"	25.7

4.2.2. Social Environment Questions

Below is the wording of questions indicated to be used for constructing the relevant variables in Table 4.3. Frequencies of each category are documented in Table 4.3.

Job satisfaction: “On the whole, how satisfied are you with the work you do -- would you say you are very satisfied, moderately satisfied, a little dissatisfied, or very dissatisfied?”

Partyid: “Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent or what?”

Voted: “In 2016, you remember that Hillary Clinton ran for President on the Democratic ticket against Donald Trump for the Republicans. Do you remember for sure whether or not you voted in that election? Did you vote for Hilary Clinton or Donald Trump?”

Political views: “We hear a lot of talk these days about liberals and conservatives. I am going to show you a seven-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale?”

Finance: “During the last few years, has your financial situation been getting better, or worse, or has it stayed the same?”

Family income: “Compared with American families in general, would you say your family income is far below average, below average, average, above average, or far above average?”

Life, news, and TV hours variables had to be dropped from the analyses due to the indicated ballot reasons in sample formation above in section 4.1.1.

Table 4.4. Social Environment Variables

Variable		**Category (ref:1)**	**Frequency Percent (%)**
JobSatisfaction	1	"Very satisfied"	43.8
	2	"Moderately satisfied"	40.9
	3	"a little dissatisfied"	11.2
	4	"Very dissatisfied"	4.1
Partyid	1	"Republican and others"	33.7
	2	"independents"	20.4
	3	"democrats"	45.9
Voted	1	"Clinton"	54.6
	2	"Trump"	37.5
	3	"Other candidate"	6.1
	4	"didnot vote"	1.8
PoliticalViews	1	"liberal"	33.3
	2	"moderate"	34.7
	3	"conservative"	32
Finance	1	"getting better"	40.4
	2	"getting worse"	20.2
	3	"stayed the same"	39.5
FamilyIncome	1	"far below average"	7.1
	2	"below average"	24.3
	3	"average"	39.9
	4	"above average"	24.9
	5	"far about average"	3.8
Life	1	"exciting"	36
	2	"pretty routine"	58.9
	3	"dull"	5.1
News	1	"everyday"	23.8
	2	"a few times a week"	13.4
	3	"once a week"	8.9
	4	"less than once a week"	16.8
	5	"never"	37.1

4.2.3. Confidence Questions

Below is the wording of questions indicated to be used for constructing the relevant variables in Table 4.5. Frequencies of each category are documented in Table 4.5.

Questions of Confidence:

"I am going to name some institutions in this country. As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them? …… (major companies/ organized religion/ education/ executive branch of the federal government/ organized labor/ press/ medicine/ TV/ US Supreme Court/ scientific community/ congress/ military/ banks and financial organizations)

Response options*:* 1. great deal, 2. only some, 3. hardly any"

Ccompanies

Creligion, etc……

Table 4.5. Confidence Variables

Question: "I am going to name some institutions in this country. As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them?....." Fill in the variables below.

Variables		**Frequency**	**Percent**	**(%)**
		1:great deal	**2:only some**	**3:hardly any**
Ccompanies	"Major companies"	16.9	61.3	21.8
Creligion	"Organized religion"	14.9	52	33.2
Ceducation	"Education"	16.7	61.4	21.9
Cfedgov	"Executive branch of fed government"	12.6	42.9	42.5
Clabor	"Organized labor"	11.2	62.5	26.3
Cpress	"Press"	11.5	41.2	47.3
Cmedicine	"Medicine"	40.2	50.6	9.2
Ctv	"TV"	7.8	50.4	41.8
CUScourt	"US Supreme Court"	25.9	54	20.1
Cscience	“Scientific Community”	50.4	43	6.6
Ccongress	“Congress”	5.3	40.9	53.8
Cmilitary	“Military”	47.2	43.2	9.6
Cfinorg	“Banks and financial organizations”	18.1	59.8	22

Answer categories 1: "great deal"
2: "only some"
3: "hardly any"

4.3. Statistical Methods

As for this section, the aim is to find out what factors can be effective in enabling data linkage. For that, the GSS data is used as a case study to investigate those factors determined. GSS provides a good source of questionnaires in three groups identified through the findings in the first chapter of this thesis. Therefore, certain characteristics discussed and decided in the survey are categorized into groups as demographic variables, social environment variables, and confidence variables. The variables within those groups are explained and the frequency distributions are observed in section 4.2.

The dependent variable is consent for data linkage, the binary variable. Because there are two response options, yes for approval or no for disapproval. The category level of consent question is one if the respondent does not agree to data linkage. It is zero, which is the reference level if the respondent gives consent for data linkage. Each of those three groups included predictors of logistic regressions.

Consent variable: **→ 0, give consent for data linkage,**
→ 1, do not give consent for data linkage

First, each independent variable is analyzed for their responses to the consent question whether there are significant differences between categories or not. Contingency table results in Chapter 6 give the distribution of each category for the frequencies of observations in that combination such as the number of males who did consent or the number of people who are very satisfied with their job and did not consent, etc. The Chi-square test implemented in this stage gives an idea of the existence of differences or not between categories.

Response, i.e. dependent variable, consent has two possible outcomes, binary. The odds for an event is p/(1-p) where p is the occurrence, probability of an event. In binary logistic regressions, odds of success are calculated according to the formula 10.

$$\frac{p}{(1-p)} = exp\,(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k), \qquad (10)$$

where k = 1, 2, …. k

The natural logarithm of odds is a linear function of the X variables, i.e. logit transformation of success measurement (formula 11)

$$Z = log\,(\frac{p}{(1-p)}) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \qquad (11)$$

Similarly in ordinary linear regressions, β_1 is the average change in Z per one unit increase in X_1, controlling for the other predictors. Differently, changes in log odds of the dependent variable, not changes in the dependent variable directly are interpreted.

The results of the statistical analyses for the micro part of this thesis explained in this section are discussed in Chapter 6.

CHAPTER 5. RESULTS OF MACRO APPROACH: SYSTEMATIC REVIEW OF CONSENT BEHAVIOR DETERMINANTS

5.1. Systematic Review Findings

5.1.1. Surveys

For the purposes of this research, within the systematic review steps, there are inclusion and exclusion criteria for what is covered in the dataset, explained in section 3.1.1. As a result of all these evaluations, 45 surveys are eligible for the analyses to answer the identified questions. There is at least one consent rate for each of those 45 surveys. The contribution of 45 surveys in statistical analysis is 128 consent rates. Table 3.1 provides a list of these surveys when and in which country they were implemented. Some surveys do not have information on consent rates for certain administrations or requests. They had to be taken out of this research, in other words, they were excluded because the consent rate is the dependent variable that is intended to be explained. On some occasions, some characteristics` information was able to be retrieved by accessing other resources including contact with the organization in combination.

5.1.2. Descriptives of Constructed Variables

Survey Consent and Response Rates. Survey response and consent rates are the only continuous variables. Table 5.1. gives the descriptive statistics of those two variables.

The average response rate of surveys in the data set constructed is 64.1 percent. And the corresponding consent rate in the same surveys is 74.8 percent.

A Pearson product-moment correlation coefficient was computed to assess the relationship between survey consent and response rate. There was a weak negative correlation between the two variables, $r = -0.1346$, $N = 128$; however, the relationship was not significant ($p = 0.1298$). The survey consent rates are not associated with the survey response rates.

Table 5.1. Descriptive Statistics for Survey Consent and Response Rate

	Size	Minimum	1st Quartile	Median	Mean	2nd Quartile	Maximum
SRR	128	9.23	48.98	72.95	64.17	84.4	96
SCR	128	15.4	59.62	84.45	74.87	95.53	100

Note: SRR: Survey Response Rate, SCR: Survey Consent Rate

Table 5.2. shows the distribution of survey category variables in categories. Those variables are discrete. Their frequencies and average consent rates for each category are presented in this table.

Table 5.2. Descriptives of Constructed Variables

Variables	Categories	Category Definitions	Counts	Percent (%)	Average Consent Rate (%)
Survey Topic	1	Other (ref.)	17	13.3	79
	2	Economic	33	25.8	67
	3	Health	78	60.9	78.9
Survey Country	1	Australia, NZ, UK (ref.)	31	24.2	73.1
	2	Us, Canada	75	58.6	75.7
	3	Germany	22	17.2	77.7
Survey Year	1	Year>=2010 (ref.)	54	42.2	80.3
	2	2001<Year<2009	44	34.4	70.9
	3	Year <=2000	30	23.4	75.4
Survey Type	1	Cross Sectional (ref.)	79	61.7	77
	2	Panel, Longitudinal, Cohort	49	38.3	71.9
Survey Mode	1	Self-Administered (ref.)	26	20.3	68
	2	CATI only	53	41.4	86.2
	3	Other (Dual, Sequential)	10	7.8	77.5
	4	CAPI only	39	30.5	62.9
Survey Target Population Age	1	All Age Groups (ref.)	44	34.4	61.8
	2	Age<18	46	35.9	89.1
	3	18<=Age	38	29.7	76.1
Survey Sponsor	1	Nongovernment (ref.)	43	33.6	67.2
	2	Government	85	66.4	79.1

Table 5.2. Descriptives of Constructed Variables Continued

Survey Consent Request Type	1	Implicit/Passive/Opt-out (ref.)	20	15.6	52.9
	2	Explicit/Active/Opt-in	108	84.4	79.6
Linkage Request Placement	1	at the End (ref.)	84	67.2	68.6
	2	at the Middle &Before Middle	41	32.8	87.4
Linkage Request Records	1	Other (ref.)	31	24.2	64.7
	2	Economic	25	19.5	76.8
	3	Health	72	56.3	79.5
Consent Wording	1	Framing (ref.)	56	47.5	79.4
	2	Neutral	62	52.5	69.6
Consent Signature and IdentityInformation	1	Signature and Idinfo (ref.)	5	4.1	85.4
	2	Signature or Idinfo	49	40.2	62.8
	3	None Required	68	55.7	81.7

Results of Boxplots.

The mean averages of each variable category are also investigated by using boxplots in Appendix A. Boxplots help to give a visual summary of descriptive distribution first by comparing averages of each category of the variable. Secondly, whiskers in the figures give the variability of category in interest. Third, skewness inside the boxes informs about the asymmetry of the category if there is one. The distribution of consent rate at each category of variables shows the most remarkable differences between consent rate averages of survey mode, consent request type, and linkage request placement. Except for survey country and linkage request placement, other variables have outliers. Some of the categories of some variables, survey type, country, sponsor, mode, linkage request placement, and consent request type are particularly skewed. In the next step, the Kruskal-Wallis test examines these significances.

Results of Kruskal Wallis test.

All the survey variables constructed except survey country and survey year show significant differences on average consent rates between their categories.

Table 5.3. Kruskal Wallis Test Results of Constructed Variables

Variable	df	p-value
Survey Topic	2	0.007098
Survey Country	2	0.3044
Survey Year	2	0.2437
Survey Type	1	5.109E-07
Survey Mode	3	0.00001222
Survey Target Population Age	2	0.003814
Survey Sponsor	1	0.001195
Survey Consent Request Type	1	0.0005241
Linkage Request Placement	1	0.00002162
Linkage Request Records	2	0.001591
Consent Wording	1	0.006075
Consent Signature and IdentityInformation	2	0.001124

****p<0.001, **p<0.01, *p<0.05 significance levels*

5.1.3. Variance Inflation Factor (VIF) Results

There is no multicollinearity between constructed survey variables found.

Table 5.4. Multicollinearity Check of Model Variables using VIF

Independent Variables	VIF
Survey Topic	1.73408
Survey Country	2.39299
Survey Year	1.36098
Survey Type	1.60921
Survey Mode	1.44845
Survey Target Population Age	1.54626
Survey Sponsor	1.59963
Survey Response Rate	2.42670
Survey Consent Request Type	1.81700
Linkage Request Placement	1.30544
Linkage Request Records	2.02966
Consent Wording	1.95967
Consent Signature and Id Info	1.64336

5.1.4. Regression Analysis

There are linear regressions fit for predicting consent rate as a function of variables identified. The significance and direction of the relationship of those variables are shown below (Table 5.1).

Survey Topic, Survey Country, Survey Year, Survey Type, and Survey Target Population Age. These are explanatory variables in the models that will be tested to see if each is related to the consent rate. It is expected that surveys with the health topic as the linkage request record will have a higher linkage consent rate compared to the other records asked for in the survey. However, both the survey topic and linkage request record are not statistically significant in explaining the data linkage consent rate. On the other hand, survey type and age of the target population have some significance in the regression model explaining consent rates. There are 49 longitudinal and 79 cross-sectional surveys in the dataset. The average consent rate for longitudinal surveys is 71.9% and 77% for

cross-sectional surveys. Longitudinal surveys have a higher consent rate than cross-sectional surveys, likely due to the trust that is maintained over a longer period of time. There are 29.7% of surveys specifically targeting 18 years and older; 35.5% of surveys targeting under 18 years; 34.4% of surveys targeting all ages. Surveys targeting younger populations are significantly associated with an average 15.8 increase in consent rates compared to those targeting older populations, showing a higher consent linkage approval in younger populations than in older populations. Country and year of survey are the variables most likely to be influenced by multiple other factors.

Survey Mode. This is searched under the categories of CATI only, CAPI only, dual & sequential, and the reference category of self-administered surveys. There are 39 CAPI, 53 CATI, 26 self-administered, and 10 dual and sequential mode surveys in the data with an average consent rate of CAPI, 62.9%; CATI, 86.2%; self-administered, 68%; and other, 77.5%. CATI surveys are significantly associated with an average increase of 18.2. In summary, CATI mode surveys have, on average, higher consent rates than CAPI compared to self-administered interviews.

Survey Sponsor. There are 85 government-sponsored surveys with an average consent rate of 79.1% and 43 non-government surveys with an average consent rate of 67.2%. On average, government-sponsored surveys have a 13.7 times higher consent rate than non-government surveys.

Survey Consent Request Type. There are 20 surveys with an implicit consent request type with an average consent rate of 52.9%. 108 surveys with an explicit consent request type with an average consent rate of 79.6%. Surveys with explicit consent request types are significantly associated with an average 26.9 times increase in consent rates compared to implicit ones.

Linkage Request Placement. There are 41 surveys with linkage request questions placed before the consent questions with an average consent rate of 87.4%; 84 surveys with linkage request questions placed at the end with an average consent rate of 68.6%. The regression model showed that surveys with the linkage request placed at the beginning

or before the consent question had a significant average 10 times increase in consent rates compared to surveys with the linkage request placed at the end.

Survey Response Rate. The consent rate and the survey response rate have a Spearman correlation coefficient of -0.15, indicating that there is almost no relationship between these two variables. The Pearson correlation was checked for the two numerical variables, the survey response rate and the consent rate, and gave similar results (section 5.1.2). As seen in Table 5.1, the regression coefficient of the response rate is -0.32 with a standard error of 0.13, which also means that there is a negative relationship between the response rate and the consent rate.

The constructed variables were also analyzed using a stepwise approach. As explained in Table 5.5., the same variables are diagnosed as significant at the same level of significance. Thus, this diagnosis gives an assurance of the reliability of the results found.

Table 5.5. Multiple and Stepwise Regression Analysis Results Comparison

Variables	Categories	Multiple	Regression	Stepwise	Regression
		Estimate	Std. Error	Estimate	Std. Error
Survey Topic	1: Other (ref.)				
	2: Economic	-3,94	7,75		
	3: Health	2,68	5,99		
Survey Country	1:Australia+NZ+UK (ref.)				
	2:US+Canada	1,39	6,52	0,87	6,08
	3: Germany	-11,67	11,19	-12.70	6,90
Survey Year	1: Year>=2010 (ref.)				
	2: 2000<Year<2010	0,59	4,48		
	3: Year<=2000	-1,5	5,26		
Survey Type	1: Cross Sectional (ref.)				
	2: Panel, Longitudinal, Cohort	15,28**	4,59	14,72**	4,41
Survey Mode	1: Other (dual, sequential) (ref.)				
	2: CATI only	18,21**	6	17,28**	5,68
	3: Self-Administered	1,59	7,62	-0.2	7,03
	4: CAPI only	1,07	5,87	0,19	5,46
Survey Target Population Age	1: All Age Groups (ref.)				
	2: Age<18	15,78**	5,25	16,14**	5,02
	3: 18<=Age	0,12	5,23	0,59	5
Survey Sponsor	1: Nongovernment (ref.)				
	2: Government	13,68**	4,69	15,06***	3,97
Survey Response Rate		-0,32*	0,13	-0,31**	0,11
Survey Consent Request Type	1: Implicit (ref.)				
	2: Explicit	26,93***	6,93	29,50***	5,96
Linkage Request Placement	1: At the End (ref.)				
	2:At the Middle &Before Consent	9,95*	3,91	10,52**	3,70

Table 5.5. Multiple and Stepwise Regression Analysis Results Comparison Continued

Linkage Request Records	1: Other (ref.)				
	2: Employment	4,83	8,44		
	3: Health	0,02	6,03		
Consent Wording	1: Framing (ref.)				
	2: Neutral	8,02	5,22	7,47	4,53
Consent Signature and Id Info	1: Signature and Idinfo (ref.)				
	2: Signature or Idinfo	-12,43	9,38	-11.11	8,59
	3: None Required	-6,23	9,74	-5.19	8,89

****p<0.001, **p<0.01, *p<0.05 significance levels*

Multiple R-squared results indicate that 66 percent of the variation in consent rates can be explained by the independent variables of the models in both multiple and stepwise regression (Table 5.6.).

Table 5.6. R_Squared Results of Models

	Multiple R_Squared	**Adjusted R_Squared**	**P value**
Multiple Regression	0.6678	0.5904	< 1.661e-15***
Stepwise Regression	0.6635	0.6079	< 2.2e-16***

CHAPTER 6. RESULTS OF MICRO APPROACH: INDIVIDUAL CONSENT BEHAVIOR IN THE GENERAL SOCIAL SURVEY

6.1. The General Social Survey Findings

The relevant questions in the GSS are grouped into three sets of variables called demographics, social environment, and trust variables. Logistic regressions are run separately for each of these sets and a full final model. The statistical results for each group are discussed in the following sections. Section 6.2. presents findings on the demographic variables of the GSS. Section 6.3. presents results on the social environment variables. Section 6.4. presents findings on trust variables. Section 6.5. presents the results of the full model with all selected variables. It explains some descriptive statistics and the results of the corresponding models that have the best statistical fit. The results of the multicollinearity tests, VIFs, are presented in the tables and interpreted for the next steps. The results of the binary logistic regressions with consent rate as the dependent variable are presented and discussed in the tables below. The models predict the number of respondents who do not consent. In other words, the models discuss the determinants of nonconsent to data linkage.

6.2. Set of Demographic Background Variables

In this section, demographic background group variables, descriptives, models, and diagnostics are interpreted statistically.

6.2.1. Results of Descriptives of Demographic Background Variables

Chi-square test results indicate that there are differences in the categories of gender and marital status once cross-tabulated with consenters and nonconsenters (Table 6.1.).

Table 6.1. Demographic Background Variables Chi-Square Test Results

Variable	**Category**	**CR[1]**	**NCR[2]**	**df**	**p-values**
Sex	"male"	733	816	1	0.0008541**
	"female"	798	1120		
Marital	"married"	718	1017	2	0.005543*
	"Not married"	412	491		
	"Never married"	400	439		
Race	"white"	1208	1449	2	0.09487
	"black"	177	222		
	"other"	127	204		
Birthcountry	"yes"	1373	1712	1	
	"no"	156	221		
Degree	"High school and below"	645	875	1	0.06978
	"Above high school"	886	1058		
Income	"Under 25K"	303	311	1	0.05997
	"25K and above"	1132	1382		
Workself	"self-employed"	164	217	1	0.6925
	"Employed by someone"	1334	1680		
Interviewregion	"West Coast"	359	454	2	0.5065
	"Midwest"	370	504		
	"East Coast"	803	993		
Health	"excellent"	344	386	3	0.2455
	"good"	844	1126		
	"fair"	285	368		
	"poor"	59	70		
Children	"none"	502	535	2	0.004703**
	"One or two children"	656	893		
	"three or more children"	366	497		

[1]CR: number of respondents who consent [2]NCR: number of respondents who do not consent

The VIF is analyzed to check if there is any collinearity existing in the models. More details about the methodology of VIF are explained in section 5.1.4.

As Table 6.2. shows there is no significant multicollinearity between model variables.

Table 6.2. VIFs of Demographic Model Variables

Independent Variables	VIF
Age	1.08609
Sex	1.030239
Marital	1.147331
Race	1.086374
Birthcountry	1.149247
Degree	1.100339
Income	1.148342
Workself	1.013152
Interviewregion	1.018549
Health	1.026555
Children	1.088747

6.2.2. Results of Logistic Regressions

The odds of not consenting to data linkage are multiplied by a factor of the corresponding odds ratio, individually for each demographic variable, all else being equal. However, none of these demographic variables are found to be significant at the 5% significance level (Table 6.3.).

Table 6.3. Binary Logistic Regressions of Demographic Variables in the GSS

	Odds Ratio	**Lower CI[1]**	**Upper CI[1]**
Age	0.997	0.993	1.002
Sex ref:"male"			
"female"	1.144	0.966	1.354
Marital ref:"married"			
notmarried	0.822	0.666	1.016
never married	0.786	0.605	1.022
Race ref: "white"			
"black"	0.925	0.709	1.206
"other"	1.032	0.736	1.447
Birthcountry ref: "US born"			
not US born	1.078	0.792	1.466
Degree ref: "highschool <="			
"> high school"	0.960	0.805	1.145
Income ref: "< 25K"			
">=25K"	1.153	0.906	1.468
Workself ref: "self employed"			
"employed by someone"	0.905	0.694	1.179
Interview Region ref: "West Coast"			
"Midwest"	1.071	0.840	1.366
"East Coast"	0.910	0.735	1.128
Health ref: "excellent"			
"good"	1.196	0.969	1.475
"fair"	1.222	0.932	1.601
"poor"	1.379	0.851	2.236
Children ref: "no child"			
"one or two children"	1.056	0.838	1.331
"three or more children"	1.032	0.795	1.339

[1]*CI: Confidence Interval*

****p<0.001, **p<0.01, *p<0.05 significance levels*

6.3. Set of Social Environment Variables

In this section, social environment group variables, descriptives, models, and diagnostics are interpreted statistically.

6.3.1 Results of Descriptives of Social Environment Variables

Chi-square test results indicate that there are differences in the categories of party identity, voting, and political views cross-tabulated with consenters and nonconsenters (Table 6.4.).

Table 6.4. Social Environment Variables Chi-Square Test Results

Variable	Category	CR	NCR	df	P_values
JobSatisfaction	"Very satisfied"	436	573	3	0.5586
	"Moderately satisfied"	433	534		
	"a little dissatisfied"	113	169		
	"Very dissatisfied"	41	51		
Partyid	"Republican and others"	456	678	2	2.15E-12
	"independents"	237	436		
	"democrats"	836	829		
Voted	"Clinton"	691	693	3	0.000002413
	"Trump"	352	543		
	"Other candidate"	71	79		
	"didnot vote"	13	29		
PoliticalViews	"liberal"	652	739	2	0.01958
	"moderate"	310	442		
	"conservative"	561	452		
Finance	"getting better"	651	761	2	0.09943
	"getting worse"	306	394		
	"stayed the same"	573	789		
FamilyIncome	"far below average"	119	128	4	1.69E-05
	"below average"	370	459		
	"average"	529	819		
	"above average"	436	480		
	"far about average"	76	59		

As Table 6.5. indicates there is no significant multicollinearity between model variables.

Table 6.5. VIFs of Socio Environment Variables

Independent Variables	VIF
JobSatisfaction	1.010484
Partyid	1.452222
Voted	1.208937
PoliticalViews	1.164664
Finance	1.053685
FamilyIncome	1.033247

6.3.2. Results of Logistic Regressions

The odds of not consenting to data linkage are 2.43 times higher for Independents than for Republicans and other party members, which is also significant at the 5% level. The odds of not consenting to data linkage are multiplied by a factor of 1.99 for each unit increase in moderate political views, all else being equal, and is statistically significant. The odds of not consenting to data linking are multiplied by a factor of 1.72 for each one-unit increase in conservative political views, which is also statistically significant. However, the odds of participants with moderate political views are more significant than those of conservatives; conservatives seem to be relatively better at consenting (Table 6.6.).

Table 6.6. Binary Logistic Regressions of Social Environment Variables in the GSS

	Odds Ratio	**Lower CI[1]**	**Upper CI[1]**
JobSatisfaction ref: "very satisfied"			
"moderately satisfied"	0.966	0.775	1.206
"a little dissatisfied"	1.170	0.828	1.654
"very dissatisfied"	0.801	0.446	1.437
Partyid ref: "republican and others"			
"independents"	2.426**	1.522	3.867
"democrats"	0.906	0.616	1.332
Voted ref: "Clinton"			
"Trump"	1.109	0.757	1.625
"Other candidate"	0.936	0.608	1.442
"didnot vote"	2.376	0.851	6.630
Political Views ref: "liberal"			
"moderate"	1.989**	1.454	2.723
"conservative"	1.719*	1.279	2.314
Finance ref: "getting better"			
"getting worse"	0.899	0.654	1.236
"stayed the same"	1.103	0.875	1.390
Family income "far below average"			
"below average"	1.336	0.709	2.520
"average"	1.249	0.670	2.327
"above average"	1.018	0.539	1.922
"far about average"	0.859	0.399	1.850

[1]*CI: Confidence Interval*

****p<0.001, **p<0.01, *p<0.05 significance levels*

6.4. Set of Confidence Variables

In this section, confidence variables, descriptives, models, and diagnostics are interpreted statistically.

6.4.1 Results of Descriptives of Confidence Variables

Chi-square test results indicate that there are differences in the categories of all confidence variables excluding US Court and military cross-tabulated with consenters and nonconsenters (Table 6.7.).

Table 6.7. Confidence Variables Chi-Square Test Results

Variable	Category	CR	NCR	df	p-values
Ccompanies	“great deal”	188	198	2	0.0143
	“only some”	577	819		
	“hardly any”	232	269		
Creligion	“great deal”	155	168	2	0.01209
	“only some”	482	696		
	“hardly any”	362	414		
Ceducation	“great deal”	193	180	2	0.002463
	“only some”	604	812		
	“hardly any”	202	292		
Cfedgov	“great deal”	161	125	2	7.06E-06
	“only some”	426	547		
	“hardly any”	411	613		
Clabor	“great deal”	142	110	2	8.25E-05
	“only some”	602	813		
	“hardly any”	250	358		
Cpress	“great deal”	156	124	2	9.22E-06
	“only some”	421	526		
	“hardly any”	419	634		

Table 6.7. Confidence Variables Chi-Square Test Results Continued

Cmedicine	"great deal"	479	466	2	1.45E-07
	"only some"	44	692		
	"hardly any"	82	129		
CUScourt	"great deal"	290	316	2	0.05192
	"only some"	515	705		
	"hardly any"	193	267		
Cscience	"great deal"	615	585	2	1.50E-13
	"only some"	336	605		
	"hardly any"	47	96		
Ccongress	"great deal"	615	585	2	0.01857
	"only some"	336	605		
	"hardly any"	47	96		
Cmilitary	"great deal"	469	602	2	0.1249
	"only some"	413	566		
	"hardly any"	113	113		
Cfinorg	"great deal"	194	219	2	0.0105
	"only some"	550	789		
	"hardly any"	255	279		

****p<0.001, **p<0.01, *p<0.05 significance levels*

As Table 6.8. shows that there is no significant multicollinearity between confidence model variables.

Table 6.8. VIFs of Confidence Variables

Independent Variables	VIF
Ccompanies	1.143843
Creligion	1.109915
Ceducation	1.12363
Cfedgov	1.161735
Clabor	1.112425
Cpress	1.199022
Cmedicine	1.172182
Ctv	1.162949
CUScourt	1.146644
Cscience	1.174576
Ccongress	1.168994
Cmilitary	1.112357
Cfinorg	1.161765

6.4.2. Results of Logistic Regressions

The odds of not consenting to data linkage for respondents with only some trust in religion are 1.41 times greater than the odds for respondents with a great deal of trust at the 5% significance level. Similarly, the odds of not consenting to data linkage for respondents with only some trust in education and medicine are 1.41 and 1.31 times greater, respectively, than the odds for respondents with a great deal of trust at the 5% significance level. Among all trust variables, trust in science is highly significant, and the odds of not consenting to data linkage are multiplied by a factor of 1.51 and 2.76 for each unit increase in only some and hardly any trust, respectively (Table 6.9.).

Table 6.9. Binary Logistic Regressions of Confidence Variables in the GSS

	Odds Ratio	**Lower CI[1]**	**Upper CI[1]**
Ccompanies ref: "great deal"			
only some	0.982	0.707	1.364
hardly any	0.677	0.455	1.008
Creligion ref: "great deal"			
"only some"	1.410*	1.005	1.981
"hardly any"	1.397	0.959	2.037
Ceducation ref: "great deal"			
"only some"	1.408*	1.019	1.947
"hardly any"	1.271	0.854	1.892
Cfedgov ref: "great deal"			
"only some"	1.297	0.892	1.886
"hardly any"	1.438	0.964	2.145
Clabor ref: "great deal"			
"only some"	1.164	0.794	1.705
"hardly any"	1.170	0.768	1.784
Cpress ref: "great deal"			
"only some"	1.238	0.850	1.803
"hardly any"	1.179	0.773	1.801
Cmedicine ref: "great deal"			
"only some"	1.309*	1.003	1.708
"hardly any"	1.172	0.748	1.837
Ctv ref: "great deal"			
"only some"	0.733	0.473	1.137
"hardly any"	0.914	0.565	1.478
CUScourt ref: "great deal"			
"only some"	1.041	0.790	1.372
"hardly any"	0.865	0.593	1.261

Table 6.9. Binary Logistic Regressions of Confidence Variables in the GSS Continued

Cscience ref: "great deal"			
"only some"	1.511**	1.164	1.963
"hardly any"	2.76***	1.634	4.662
Ccongress ref: "great deal"			
"only some"	0.851	0.501	1.446
"hardly any"	0.960	0.552	1.669
Cmilitary ref: "great deal"			
"only some"	0.912	0.711	1.169
"hardly any"	0.666*	0.448	0.990
Cfinorg ref: "great deal"			
"only some"	0.869	0.619	1.221
"hardly any"	0.865	0.572	1.309

[1]*CI: Confidence Interval*

****p<0.001, **p<0.01, *p<0.05 significance levels*

6.5. Results of Full Model

All sets are combined to discuss a full model for consent behavior.

The variables selected for the model do not have any multicollinearity as indicated in the analysis of VIF in Table 6.10.

Table 6.10. VIFs of All Selected Variables

Independent Variables	VIF
Age	1.409847
Sex	1.128747
Marital	1.227133
Race	1.228362
Birthcountry	1.284011
Degree	1.253034
Income	1.261305
Workself	1.079644
Interviewregion	1.083407
Health	1.090237
Children	1.188186
JobSatisfaction	1.088317
Partyid	1.653263
Voted	1.322217
PoliticalViews	1.316872
Finance	1.11811
FamilyIncome	1.170555
Ccompanies	1.218133
Creligion	1.203359
Ceducation	1.155025
Cfedgov	1.253211
Clabor	1.164401
Cpress	1.311368
Cmedicine	1.246034
Ctv	1.234169
CUScourt	1.200348
Cscience	1.286768
Ccongress	1.216282
Cmilitary	1.201433
Cfinorg	1.247604

The final model examines all predictors in a single model. The odds of not consenting to data linkage are 3.89 and 4.59 times higher, respectively, for independents who did not vote than for Republicans and others who voted for Clinton. The odds of not consenting to data linkage are multiplied by a factor of 1.82 for each one-unit increase in moderate political views achieved, all else being equal, and it is statistically significant. Trust in science is highly significant, the odds of not consenting to data linkage are multiplied by a factor of 1.56 and 4.84 for each one-unit increase in only some and almost none. In addition, trust in the military is statistically significant at a 5% level, and the odds of not consenting to data linkage are multiplied by a factor of 0.54 for each one-unit increase in confidence in the military only some relative to a great deal. Similarly, it is multiplied by a factor of 0.34 for each one-unit increase in confidence in the military hardly any to a great deal. The McFadden pseudo-R-squared explains 71 percent of the model with these independent variables.

Table 6.11. Binary Logistic Regressions of all Variables selected in the GSS

	Odds Ratio	**Lower CI[1]**	**Upper CI[1]**
Age	0.995	0.979	1.011
Sex ref:male			
female	1.246	0.873	1.780
Marital ref:married			
notmarried	0.847	0.542	1.323
never married	0.826	0.490	1.390
Race ref: white			
black	1.006	0.550	1.838
other	0.789	0.350	1.779
Birthcountry ref: US born			
not US born	0.634	0.309	1.304
Degree ref: highschool <=			
>highschool	0.866	0.578	1.297

Table 6.11. Binary Logistic Regressions of all Variables selected in the GSS Continued

Income ref: < 25K			
>25K	1.854	0.977	3.518
Workself ref: self employed			
employed	0.952	0.553	1.638
InterviewRegion ref: West Coast			
Midwest	0.838	0.510	1.378
East Coast	0.660	0.425	1.025
Health ref: excellent			
good	1.365	0.911	2.045
fair	1.044	0.596	1.827
poor	1.077	0.177	6.546
Children ref: no child			
one or two	1.109	0.693	1.774
three or more	1.055	0.622	1.789
JobSatisfaction ref: very satisfied			
moderately satisfied	0.862	0.593	1.253
alittle dissatisfied	0.754	0.411	1.383
very dissatisfied	1.311	0.475	3.614
Partyid ref: republican and othrs			
independents	3.898***	1.783	8.523
democrats	1.353	0.710	2.579
Voted ref: Clinton			
Trump	0.974	0.509	1.867
other candidate	1.650	0.783	3.478
did not vote	4.592*	1.018	20.709
Political Views ref:liberal			
moderate	1.822*	1.045	3.179
conservative	1.355	0.785	2.338
Finance ref: getting better			
getting worse	1.088	0.632	1.875
stayed the same	1.232	0.837	1.812

Table 6.11. Binary Logistic Regressions of all Variables selected in the GSS Continued

Family income far below average			
below average	1.654	0.477	5.730
average	1.860	0.530	6.526
above average	1.441	0.392	5.300
far about average	1.121	0.247	5.085
Ccompanies ref: great deal			
only some	1.129	0.675	1.888
hardly any	1.247	0.650	2.391
Creligion ref: great deal			
only some	1.038	0.599	1.800
hardly any	1.403	0.752	2.619
Ceducation ref: great deal			
only some	1.361	0.834	2.220
hardly any	1.195	0.640	2.232
Cfedgov ref: great deal			
only some	0.865	0.458	1.634
hardly any	1.210	0.623	2.353
Clabor ref: great deal			
only some	0.848	0.501	1.434
hardly any	0.988	0.537	1.818
Cpress ref: great deal			
only some	1.384	0.768	2.493
hardly any	1.337	0.662	2.703
Cmedicine ref: great deal			
only some	1.359	0.894	2.066
hardly any	0.678	0.327	1.408
Ctv ref: great deal			
only some	0.837	0.409	1.716
hardly any	1.023	0.470	2.230
CUScourt ref: great deal			
only some	1.095	0.719	1.669

Table 6.11. Binary Logistic Regressions of all Variables selected in the GSS Continued

hardly any	0.646	0.351	1.186
Cscience ref: great deal			
only some	1.558*	1.004	2.418
hardly any	4.843***	1.930	12.157
Ccongress ref: great deal			
only some	0.821	0.327	2.059
hardly any	0.865	0.331	2.264
Cmilitary ref: great deal			
only some	0.539**	0.363	0.801
hardly any	0.337**	0.166	0.684
Cfinorg ref: great deal			
only some	0.612	0.363	1.031
hardly any	0.792	0.403	1.553

[1]*CI: Confidence Interval*

****p<0.001, **p<0.01, *p<0.05 significance levels*

CHAPTER 7. CONCLUSIONS AND DISCUSSION

In this thesis, a systematic review was first conducted to access survey literature with consent rates for linking survey and administrative data. The literature with this type of consent was aggregated and summarized and several aspects of the consent question that affect consent rates were discovered. With what is learned in the first section, a detailed analysis of one of the surveys GSS is implemented.

The survey literature, as explained, found the importance of four factors -survey mode, survey sponsor, response rates, and linkage request placement- in explaining the consent behavior. These findings in the literature are supported by the results of this thesis.

The results of the survey mode show that respondents still prefer human interaction, although new technologies such as online web surveys limit this to some extent. Even though it is not face-to-face, having the possibility to communicate with interviewers about the survey study via CATI increases the likelihood of obtaining higher consent rates. Interviewers can help respondents overcome their lack of knowledge and provide them with the descriptions they need to make an informed decision, which can ultimately lead to higher consent rates for data linkage. In addition to the security concerns with online platforms, the need to read and understand seems to be a task that respondents are not always willing to take on. Face-to-face contact with respondents works comparatively well and clarifies the respondents' possible questions to obtain consent for linkage, similar to obtaining a response. (Jenkins, et al., 2006).

The results on sponsors, linkage request placement, consent request type, and survey type prove once again the importance of trust for participants. As is well known, there can be a variety of reasons for participating as a respondent, with government-sponsored surveys with a sense of obligation to participate in some cases being a stress factor for respondents or something else (Linsky, 1975; Goyder, 1985; De Leeuw & Heer, 2002).

Clearly stating the consent question at the beginning, maintaining a longer correspondence of more than one period as in panel surveys, and the government as the

main organization to be trusted are likely to lead to higher consent approval for linkage. Asking for consent as early as possible in panel surveys such as the HRS is observed to result in a higher rate of positive consent (Sakshaug (2021). Longitudinal surveys have higher consent rates than cross-sectional surveys. In panel surveys, when the survey organization establishes a relationship with the respondent by providing detailed information about consent and also by mentioning the period of the survey, they are already creating a kind of trust. In addition, asking for consent explicitly rather than implicitly also supports the trust mechanism. Therefore, as further research, it would be interesting to look into the interactions between some trust-related variables once a sufficient sample size is reached**.**

Getting knowledge in the first part of the thesis, which is the macro approach, the effect of trust is studied as a group of variables in the GSS at the micro approach. Specifically, trust in science, military, and voting behavior show an impact on agreement or disagreement. Respondents who have a high level of trust in the military, identify themselves as independents, and have moderate political views are more likely to consent than others. There are no significant effects and differences by demographic characteristics to influence consent behavior. Sakshaug & Kreuter's (2012) research also found a relatively small nonconsent bias for estimates of demographic variables. Demographic characteristics such as education, age, and gender differences are relatively insignificant to give a positive or negative response.

Although the literature has attempted to explain the consent issue in a framework similar to nonresponse theory, the behavioral characteristics of nonresponse and data linkage consent are not necessarily the same. The contribution of factors such as effective interviewers, advance letters, and refusal conversion efforts to unit response and the desire to consent to data linkage are predicted to be positively related (Groves, et al., 2004). Contrary to popular belief, this research shows that surveys with lower response rates have higher consent for linkage approval rates than surveys with higher response rates. Respondents may be "willing" to answer the survey questions, but they may not feel

comfortable sharing their private information by approving the linkage consent. Therefore, even though the response rate may be high, the consent rate may be lower. Similar to the findings of Sakshaug & Kreuter (2012), there is not necessarily a relationship between non-consent and non-response. Another argument would be that in a survey with a lower response rate, sample respondents are already resistant at the response stage, and the remaining respondents have a higher consent.

In summary, the results of our analysis strengthen the importance of trust built with the survey participants in every field. As Grove's Leverage Salience Theory motivates, any features associated with building any type of relationship online, incentives, etc., increase the number of consenters. They are motivated. Trusting the survey, in which they participate facilitates survey researchers to correspond and commit to the following step as consent linkage which is their confidential information to respond at the end. Thus, any survey study via the usage of all possible resources should aim to build trust with respondents in the first stage to maintain positive returns for higher consent rates.

Taking all those into account, below are the limitations of this study and our recommendations for survey designers on creating surveys, that ask consent for data linkage.

Limitations.

Typically, the results have some limitations. First, the research only in the language of English is considered in the systematic review application. There is a possibility to have surveys with consent asked in other language speaking countries as well. However, consent research, which is the result of detailed research and strategies, is relatively more common in English-speaking countries. Therefore, this is not a disadvantage. However, this study, being a systematic review with additional statistical analysis, is interpreted in the context of the aforementioned studies. It is also the largest systematic review of consent rates.

We also understand that the GSS is one of many surveys. This research can be applied to many other surveys, including some with trust questions. In this way, the results

can be observed in a larger sample of surveys to see the relationships between trust and consent to data linking. We believe that there will be further research in the future that takes advantage of our findings on this issue of consent for data linkage.

Despite these limitations, the results are valuable to researchers studying consent behavior for at least two reasons: 1) they are the first to provide a systematic review of this large sample on consent rates based on the survey literature; 2) they identify some of the key factors related to building trust with the survey organization that are associated with higher consent rates. Therefore, the results may provide some useful insights for researchers in designing their survey studies.

Recommendations.

If a survey is to be designed with a survey consent question to ask for data linkage, building up on this thesis, there would be some main recommendations to the researchers or policymakers. The first step should be to clarify steps on how to create a trustful, motivating environment for the respondents, particularly for the given topic of the survey. As leverage salience theory proves, motivational procedures interactively help to facilitate receiving informed consent, which gives clues to be a similar case for consent for data linkage. Those procedures include the consent question placement, survey mode, and interviewer effect. This thesis agrees that the placement of consent questions for data linkage closer to the beginning of the survey appeals to more positive replies of higher consent. Besides, it is witnessed that the survey mode with second parties as telephone or face-to-face interaction also adds value to positive consent. However, due to the limited access to the interviewer data, we were not able to look into factors relevant to interviewers in this thesis. If any resources of a group of sampled surveys with access to the demographics of interviewers are possible in the future research, there can be additional variables constructed to find their power to explain consent rates through statistical models. Gender, education, and age can be some of those demographics of interviewers to be listed for research. Our recommendations of a good survey with consent for data linkage questions can apply to any country including Türkiye. As expected, the

country predictor does not differentiate in explaining consent because multiple factors in countries play a role in consent decision consequently. What is most important for each country is its regulations regarding the privacy and confidentiality of its citizens' data. Accordingly, they make efforts not only to obtain consent for data linkage but also for informed consent for survey participation and health data access.

CHAPTER 8. REFERENCES

Angus, V.C., Entwistle V.A., Emslie M.J., Walker, K.A., & Andrew, J.E. (2003). The requirement for prior consent to participate on survey response rates: a population-based survey in Grampian. BMC Health Services Research. 3, 21. http://www.biomedcentral.com/1472-6963/3/21.

Baghal, T.A., Knies, G. & Burton, J. (2014). Linking administrative records to surveys: differences in the correlates to consent decisions. *Understanding Society Technical Report,* No. 2014-09, Institute for Social and Economic Research, Essex, UK.

Baghal T.A., Sloan, L., Jessop, C., Williams, M.L. & Burnap, P. (2019). Linking twitter and survey data: the impact of survey mode and demographics on consent rates across three UK studies. Social Science Computer Review, XX(X), 1-16.

Bates, N. (2005). Development and testing of informed consent questions to link survey data with administrative records. Proceedings of the Survey Research Methods Section of the American Statistical Association, 3786–3793.

Blumberg, S. J. & Cynamon, M. L., (1999). "Misreporting Medicaid Enrollment: Results of Three Studies Linking Telephone Surveys to State Administrative Records," Proceedings of the 7th Conference on Health Survey Research Methodology, DHHS Publication No. 01-1013. Hyattsville, MD.

Buckley, B., Murphy, A.W., Byrne, M., Glynn, L. (2007). Selection bias resulting from the requirement for prior consent in observational research: community cohort of people with ischemic heart disease. Heart, 93, 1116–1120.

Carter, K., Shaw, C., Hayward, M. &Blakely, T. (2010). Understanding the determinants of consent for linkage of administrative health data with a longitudinal survey. New Zealand Journal of Social Sciences, 5(2), 53-60.

Cehovin, G., Bosnjak, M. & Manfreda, K., L. (2018). Meta-analyses in survey methodology: a systematic review. Public Opinion Quarterly, 82(4), 641–660.

Cornesse, C., Bosnjak, M. (2018). Is there an association between survey characteristics and representativeness? A meta-analysis, Survey Research Methods, 12(1), 1-13.

Couper, M. & Singer, E. (2013). Informed consent for web paradata use. Survey Research Methods, 7(1), 57-67.

Daikeler, J., Bosnjak, M., Manfreda, K., L. (2019). Web survey versus other survey modes: A meta-analysis comparing response rate. Journal of Statistics and Survey Methodology, 0, 1-27.

Das, M., Couper, M., P. (2004). Optimizing opt-out consent for record linkage. Journal of Official Statistics, 30(3), 479–497.

Dormann, C.F., et al. (2013). Collinearity. A review of methods to deal with it and a simulation study evaluating their performance. Ecograpy, 36, 27-46.

De Leeuw, E., Heer, W.D. (2002). "Trends in household survey nonresponse: a longitudinal and international comparison." In Survey Nonresponse, eds. R. Groves, D. Dillman, J. Eltinge, and R. J. A. Little. New York: Wiley.

Dunn, K.M., Jordan, K., Lacey, R.J., Shapley, M. & Jinks, C. (2004). Patterns of consent in epidemiological research: evidence from over 25000 respondents. The Practice of Epidemiology, 159(11), 1087-1094.

Elevelt, A. (2021). Consent to data linkage in surveys: A descriptive review and meta-analysis. Dissertation on Smart (Phone) Surveys.

Fobia, A.C., Holzberg, J., Eggleston, C., Childs, J.H., Marlar, J., Morales, G. (2019). Attitudes towards data linkage for evidence-based policymaking. Public Opinion Quarterly, 83, Special Issue, 264-279.

Fulton, J. A. (2012). Respondent consent to use administrative data. Unpublished Ph.D. Dissertation, University of Maryland, Available at http://drum.lib.umd.edu/handle/1903/13601.

Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5(10), 3-8.

Gessendorfer, J., Beste, J., Drechsler, J. & Sakshaug, J.W. (2018). Statistical matching as a supplement to record linkage: a valuable method to tackle nonconsent bias. Journal of Official Statistics, 34(4), pp:909-933.

Goyder, J. (1985). Face-to-face interviews and mailed questionnaires: the net difference in response rate. Public Opinion Quarterly, 49(2), pp:234-252.

Groves, R.M., Cialdini, R.B. & Couper, M.P. (1992). Understanding the decision to participate in a survey. Public Opinion Quarterly, 56, 475–495.

Groves, R.M., Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. a meta-analysis. Public Opinion Quarterly, 72(2), 167-189.

Groves, R.M., Presser, S., Dipko, S. (2004). The role of topic interest in survey participation decisions. Public Opinion Quarterly, 68(1), 2-31.

Groves, R.M., et.al. (2004). Survey Methodology. John Wiley & Sons.

Groves, R., M., Singer, E. & Corning, A. (2000). Leverage-saliency theory of survey participation: description and an illustration. Public Opinion Quarterly, 64(3), 299-308.

Hair, J.F., (2014). Multivariate data analysis. Pearson new international edition, 7th ed. Upper Saddle River, NJ, Prentice Hall.

Haraldsen, G., Jones, J., Giesen, D. & Zhang, L., C. (2013). Understanding and coping with response burden. Designing and Conducting Business Surveys eds. G. Snijkers, G. Haraldsen, J. Jones, and D. K. Willimack, 219–252, Hoboken, NJ: John Wiley and Sons.

Harris, T., et.al. (2005). Linking questionnaires to primary care records: factors affecting consent in older people. Journal of Epidemiological Community Health, 59, 336-338.

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. American Sociological Review, 43, 447–462.

Huang, et. al (2007). Record linkage research and informed consent: who consents? BMC Health Services Research, 7:18.

Jäckle, A., Beninger, K., Burton, J. & Couper, M.P. (2021). Understanding data linkage consent in longitudinal surveys. Advances in Longitudinal Survey Methodology eds. Peter Lynn, Wiley Series in Probability and Statistics.

Jäckle, A., Burton, J., Couper, M. & Crossley, T. (2021). Understanding and Improving Data Linkage Consent in Surveys, 2018-2019. Data Collection. Colchester Essex.

Jäckle, A., Sala, E., Jenkins, S., P. & Lynn, P. (2004). Validation of survey data on income and employment: the ISMIE experience. ISER Working Paper, No. 2004-14. Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-14.pdf

Jenkins, S. P., Cappellari, L., Lynn, P., Jäckle, A. & Sala, E. (2006). Patterns of consent: evidence from a general household survey. Journal of the Royal Statistical Society: Series A (Statistics in Society). 169, 701–722.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. Public Opinion Quarterly, 83, 210- 235.

Kho, M.E. (2009). Written informed consent and selection bias in observational studies using medical records: systematic review. BMJ Research. https://doi.org/10.1136/bmj.b866

Kim, J.K. & Rao, J.N.K. (2012). Combining data from two independent surveys: a model-assisted approach. Biometrika 99(1), 85-100.

Kreuter, F., Muller, G. & Trappmann, M. (2010). Nonresponse and measurement error in employment research: making use of administrative data. Public Opinion Quarterly, 74, 880-906.

Kreuter, F., Sakshaug, J.W. & Tourangeau, R. (2016). The framing of the record linkage consent question. International Journal of Public Opinion Research, 28(1), 142-152.

Krosnick, J. A. (2018). Improving question design to maximize reliability and validity. In D. L. Vannette & J. A. Krosnick (Eds.), The Palgrave Handbook of Survey Research New York: Palgrave Macmillan, 95- 101.

Linsky, A. S. (1975). Stimulating responses to mailed questionnaires: A review. Public Opinion Quarterly, 39, 82-101.

Lugtig, P. (2014). Separating stayers, fast attriters, gradual attriters, and lurkers. Sociological Methods & Research. 43, 699-723.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. (2010). Preferred reporting items for systematic reviews and meta-analyses: the Prisma statement. International Journal of Surgery, 8(5), 336-341.

Olson, K. M. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. Public Opinion Quarterly, 70, 737-58.

Pascale, J. (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). US. Census Bureau Working Paper, 2011(3).

Rubin, D. B. (1976). Inference and missing data. Oxford Journals, Biometrika Trust, 63, 581-592.

Sakshaug, J. W., and Antoni, M. (2019). Evaluating the Utility of Indirectly Linked Federal Administrative Records for Nonresponse Bias Adjustment. Journal of Survey Statistics and Methodology, 7(2), 227-249.

Sakshaug, J. W., Couper, M. P., Ofstedal, B. & Weir, D.R. (2012). Linking survey and administrative records mechanisms of consent. Sociological Methods and Research, 41, 535–569.

Sakshaug, J. W. & Huber, M. (2016). An evaluation of panel nonresponse and linkage consent bias in a survey of employees in Germany. Journal of Survey Statistics and Methodology, 4, 71–93.

Sakshaug, J.W., Hülle, S., Schmucker, H., & Liebig, S. (2017). Exploring the effects of interviewer- and self-administered survey modes on record linkage consent rates and bias. Survey Research Methods, 11(2), 171-188.

Sakshaug, J, W., Schmucker, A., Kreuter, F., Couper, M. P. & Singer, E. (2019). The effect of framing and placement on linkage consent. Public Opinion Quarterly, 83(special issue), 289-308.

Sakshaug, J, W. & Vicari, B, J. (2018). Obtaining record linkage consent from establishments: the impact of question placement on consent rates and bias. Journal of Survey Statistics and Methodology, 6, 46–71.

Sakshaug, J., W. & Kreuter, F. (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, J. W. & Kreuter, F. (2012). Assessing the magnitude of non-consent biases in linked survey and administrative data. Survey Research Methods. 6, 113–122.

Sakshaug, J, W., Schmucker, A., Kreuter, F., Couper, M. P. & Singer, E. (2016). Evaluating active (opt-in) and passive (opt-out) consent bias in the transfer of federal contact data to a third-party survey agency. Journal of Survey Statistics and Methodology, 4, 382-416.
Sakshaug, J. W., Tutz, V. & Kreuter, F. (2013). Placement, wording, and interviewers: identifying correlates of consent to link survey and administrative data. Survey Research Methods, 7, 133-144.

Sakshaug, J. W., Yan, T. & Tourangeau, R. (2010b). Nonresponse error, measurement error, and mode of data collection method: tradeoffs in a multi-mode survey of sensitive and non-sensitive items. Public Opinion Quarterly. 74, 907-933.

Sala, E., Burton, J. & Knies, G. (2010). Correlates of obtaining informed consent to data linkage: respondent, interview, and interviewer characteristics. Institute for Social and Economic Research, 2010-28.

Sala, E., Knies, G, & Burton, J. (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.

Sala, E., Burton, J. & Knies, G. (2012). Correlates of obtaining informed consent to data linkage: respondent, interview, and interviewer characteristics. Sociological Methods & Research. 41, 414-439.

Shahi, R. & Vousden, C. (2005). Bias from requiring explicit consent from all participants in observational research: prospective, population-based study. BMJ. doi:10.1136/bmj.38624.397569.68

Singer, E. (1978). Informed consent: consequences for response rate and response quality in social surveys. American Sociological Review, 43(2), 144-162.

Singer, E. (1978). The effect of informed consent procedures on respondents' reactions to surveys. Consumer Research, 5(1), 49-57.

Singer, E. (1984). Public reactions to some ethical issues of social research: attitudes and behavior. Journal of Consumer Research, 11(1), 501-509.

Singer, et al., (1993). Informed consent and survey response: A summary of the empirical literature. Journal of Official Statistics, 9(2), 361-375.

Singer, et al., (1999). Differential Incentives: Beliefs About Practices, Perceptions of Equity, and Effects on Survey Participation. Journal of Official Statistics, 63(2), 251-260.

Singer, E. (2003). Exploring the meaning of consent: participation in research and beliefs about risks and benefits. Journal of Official Statistics, 19, 273–285.

Smith, T. et al (2021). General Social Survey, 2021 Cross Section Study Codebook. www.norc.org/documentation.

Smith, Tom W. and Kim, Jibum, (2013) “An Assessment of the Multi-level Integrated Database Approach,” GSS Methodological Report No. 116. Chicago: NORC, Published in The ANNALS of the American Academy of Political and Social Science, 646, 185-221.

Thornby, M., Calderwood, L., Kotecha, M., Beninger, K. & Gaia, A. (2017). Collecting multiple data linkage consents in a mixed mode survey: evidence and lessons learned from next steps. Center for Longitudinal Studies Working Paper, 2017/13. doi:10.13094/SMIF-2018-00004

Tourangeau, R., Rips, L.J. & Rasinski, K. (2000). The Psychology of Survey Response. Cambridge and New York: Cambridge University Press.

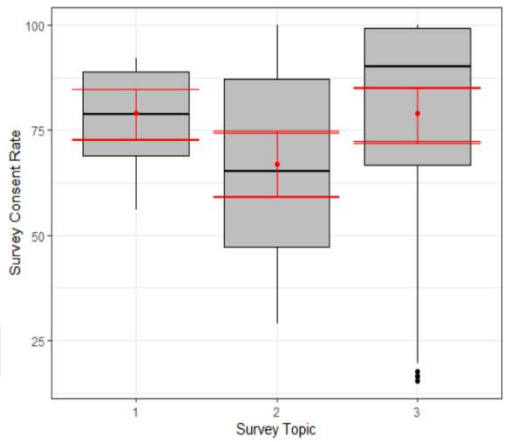
Webster, C. (1996). Hispanic and Anglo interviewer and respondent ethnicity and gender: the impact on survey response quality. Journal of Marketing Research, 33(1), 62-72.

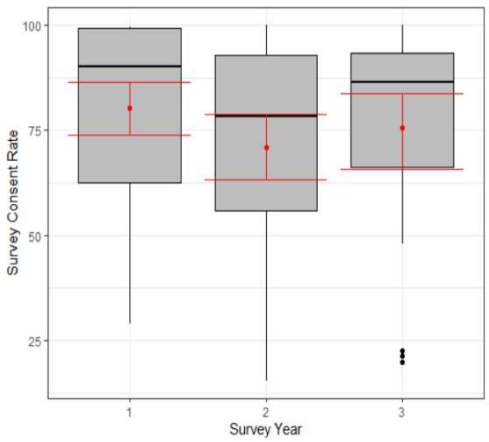
Willimack, D. K. & Snijkers, G. (2013). The business context and its implications for the survey response process. in Designing and Conducting Business Surveys, eds. G. Snijkers, G. Haraldsen, J. Jones, and D.K. Willimack, 39–82, Hoboken, NJ: John Wiley and Sons.

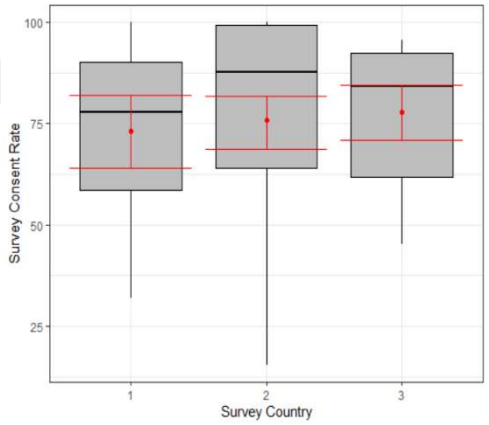
Yawn BP, et. al. (1998). The impact of requiring patient authorization for use of data in medical records research. Journal of Family Practice, 47:361-365.

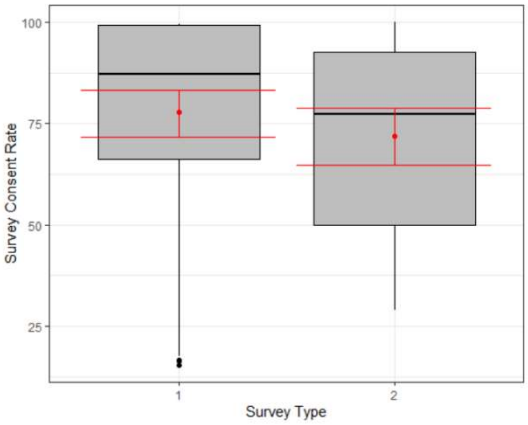
APPENDIX

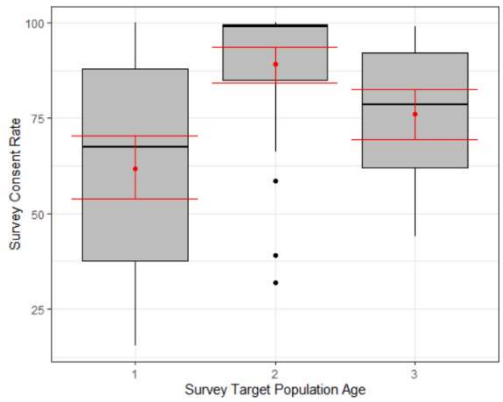
APPENDIX A. Boxplots

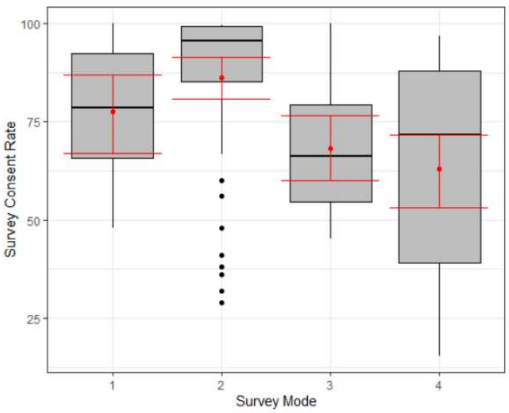


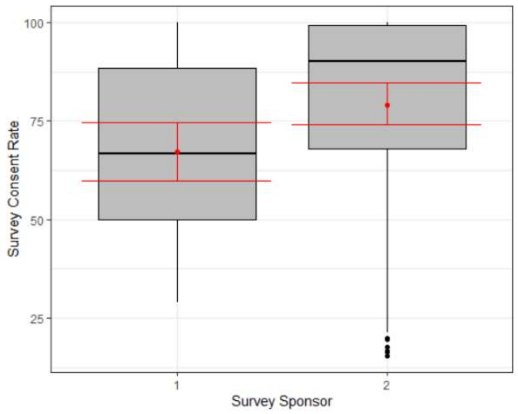


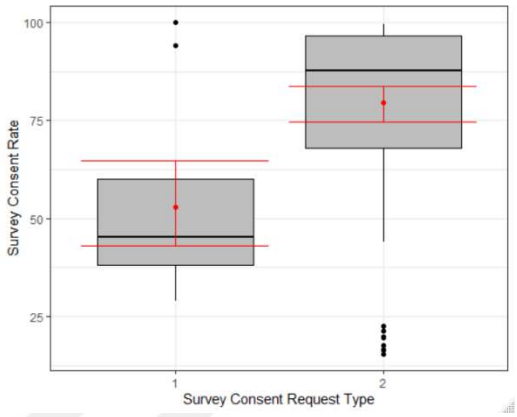


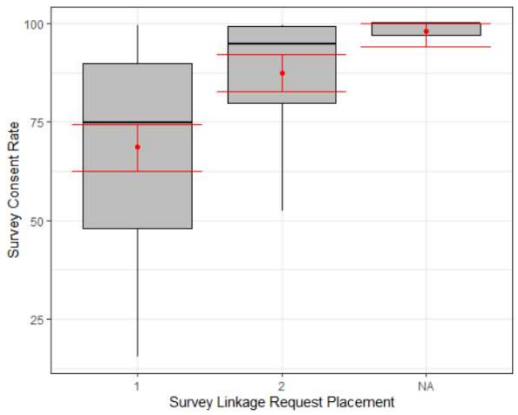


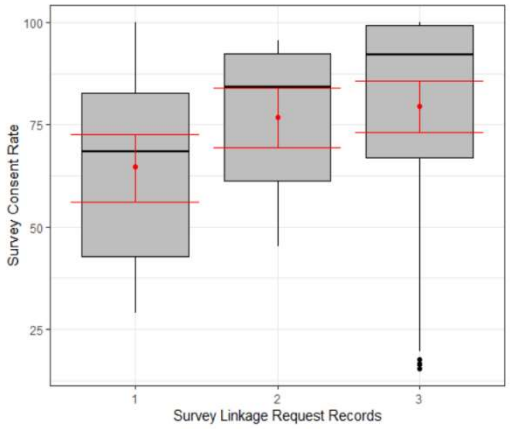


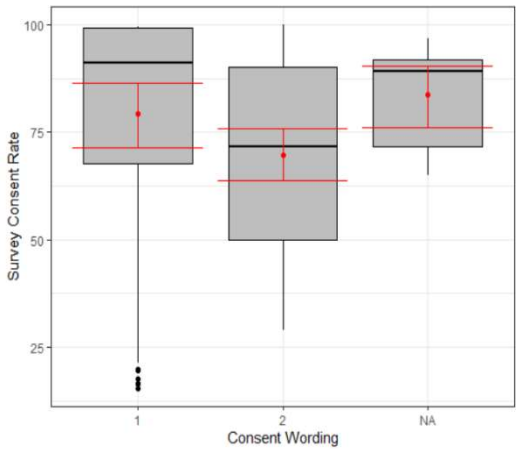


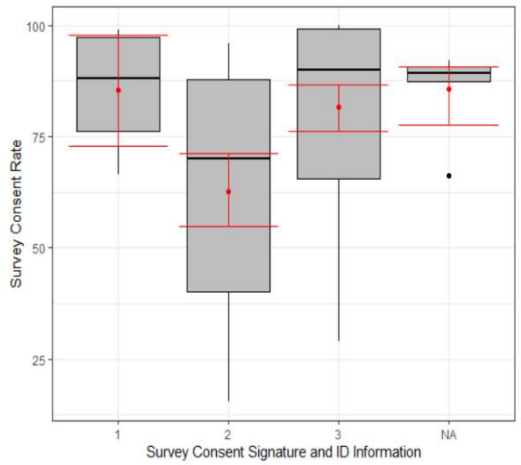

Survey Consent Rate
100
75
50
25
1
2
Survey Sponsor


Survey Consent Rate
100
75
50
25
1
2
Survey Consent Request Type


Survey Consent Rate
100
75
50
25
1
2
NA
Survey Linkage Request Placement


Survey Consent Rate
100
75
50
25
1
2
3
Survey Linkage Request Records


Survey Consent Rate
100
75
50
25
1
2
NA
Consent Wording


Survey Consent Rate
100
75
50
25
1
2
3
NA
Survey Consent Signature and ID Information

APPENDIX B. GSS Consent Information Sheet

Administrative Data Linkage Consent

One important purpose of our survey is to gather information about our society in order to explain trends in attitudes and behaviors. The GSS seeks to support the study of opinion data with more behavioral, fact-based information. To do this, we need your permission to search government records that may exist now or in the future such as Census data, unemployment or social security information, Medicaid and Medicare, health records and other sources.

We would link your interview with that information upon completion of the survey. Also, we would like to be able to continue combining these data for the foreseeable future. Each time, in the future, we will remove personally identifiable information after the linking process, and store anonymous data for the duration of our study. The best way for us to get this information without taking up any more of your time is to conduct the work in a secure data facility meeting strict Federal Information Security Management Act requirements, and we need your permission to proceed. All identifying information will be kept confidential as required by law. If at any time you change your mind, you can contact us at the toll-free number 888-837-8988.

May we try to link government records with your survey answers?

Why are you asking to link my data? Why is it important?

- Linking survey data to other existing sources of information is important because it will help us collect more information without having to request more of your time.
- By linking data, people don't have to remember information that may not always be immediately available in an interview.
- The GSS is one of the largest and oldest studies in the U.S. conducted on behalf of the National Science Foundation, you will be making a difference to understand how our society works.
- Linking data from the GSS to administrative data allows the project to analyze the answers you and others have given with different types of information captured in administrative records over time without having to ask you complete additional surveys or questionnaires. Analysis of this type of trend data is important in helping to understand behaviors in our society.
- Statistics generated with results of the General Social Survey are usually cited in the media and decision makers in order to inform their understanding of the social and financial reality in the country.

What kind of information will you be linking my data with?

- With your help, for example, we can learn more about how unemployment status, social security information, Medicaid and Medicare status, or health records and other behaviors in the future may be associated with people's decision making and attitudes today.
- With your permission, we would link survey data to sources available to government agencies, for example, health insurance coverage, government benefits such as unemployment or Social Security information, Medicare or Medicare records, housing aid, Student loans, among others.

How will my data be protected? How do I know my information will be kept safe?

- The same commitment to confidentiality that protects your survey answers also protects this additional information.
- All researchers who handle the information collected for the study must sign a confidentiality agreement that provides for criminal and civil penalties if confidentiality is breached.
- We would provide basic information that identifies you to government agencies, and request that information about you to be linked to your opinions. For example, we would coordinate with the Census Bureau's Federal Statistical Research Data Centers to use a numeric code for you to instead of any identifiable personal information. This numeric code is known as Protected Identification Key and helps combine the survey data with administrative records.
- Federal Privacy Regulations provide safeguards for privacy, security, and authorized access, such as the Federal Information Security Management Act (FISMA).
- Our study has also obtained a federal Certificate of Confidentiality from the National Institutes of Health in order to further protect your privacy. With this Certificate, we cannot be forced (for example by court order or subpoena) to disclose information that may identify you in any federal, state, local, civil, criminal, legislative, administrative, or other proceedings.

How will my data be used? Who will see my answers?

- Only a few people who work on this survey ever see any personal information. Answers that could identify you in any way are separated from your other answers.
- Survey findings are put into summary reports that contain no names or other information that identifies you.
- In other words, your name and identifying information will never be released to the public, only combined with information from others such that your confidentiality is protected.

Can I change my mind later about providing my permission?

- Yes, you can always choose not to provide consent for linking your data.
- You can do so, by calling or emailing us at 1-(888)-837-8988 or mygss@norc.org to let us know that you have changed your mind.

APPENDIX C. Additional Variables Eliminated

Initially, more variables are created, including survey period, sampling unit, survey consent mode, survey proxy response, and target population type. However, due to reasons such as missing data or accuracy, they are not included in the regression models. The definitions of these variables are as follows.

Survey Period. This variable indicates the frequency period of the survey, how often, in what time frame the survey is conducted. It is structured in categories of one-time study, usually experiments, or annual or every 2 years or other. Our question was whether or not knowing that the survey runs for more than one period would influence respondents' willingness to consent to data linkage. There are surveys where participants do not know that they will receive the same questionnaire in the next month or so until the end. Then it depends on at what point in the interview they find out. Even if they do, whether or not they are still interested in participating and giving consent for data linkage. There are those respondents who will be informed that their data have been linked from wave one unless they opt out at a later wave of the survey. The decision to withdraw their consent at a later point in the survey can be easily handled, as it is clearly explained to them in the consent question. It may also be a cross-sectional survey that requires consent from a new sample of respondents in each period. These categories are covered by the survey type categories of cross-sectional or longitudinal. The survey type is selected instead.

Survey Consent Mode. This is the method of collecting consent for data linkage authorization. Is it written, verbal, or both? Self-administered surveys are via written forms to indicate their responses to consent. On the other hand, telephone surveys CATI consent either oral or written forms sent to be signed. Since consent mode changes with the type of data collection mode of the survey, a decision was made not to include consent mode as a predictor in models.

Survey Proxy Response. This variable indicates whether the respondent is a proxy or not. However, survey information on this detail was not publicly available.

Target Population Type. Specific or rare populations are likely to differ from the rest of the population. For example, for an Internet usage survey, maintaining a population of Internet access to draw the sample is a specific one. However, for this thesis, there are no specific population surveys found. Then this option is automatically excluded.

Additional Variables

Variable	Variable Explanation	Categories
Survey Period	frequency of the survey implemented	1:1time study, 2: annual, 3: every 2 years 4: other
Survey Consent Mode	mode of asking consent for data linkage	1: written, 2: oral, 3: written and oral
Survey Proxy Response	if respondent is proxy or not	1: yes, 2:no
Target Population Type	whether rare population is studied or not	1: general, 2:specific

APPENDIX D. Original Article

Fiscaoeconomia

E-ISSN: 2564-7504

2024, Volume 8, Issue 2, 495-516

https://dergipark.org.tr/tr/pub/fsecon

Submitted/Geliş: 30.12.2023

Accepted/Kabul: 10.03.2024

Doi: 10.25295/fsecon.1412153

Research Article/Araştırma Makalesi

Investigating the Determinants that Influence Consent Behavior for Linking Survey Data with Administrative Records

Anket Verilerin İdari Kayıtlarla Birleştirilmesi İçin Onay Verme Davranışını Etkileyen Faktörlerin Araştırılması

Türknur BRAND[1], Ahmet Sinan TÜRKYILMAZ[2]

Abstract

This research investigates the determinants of consent behavior in linking survey and administrative data, filling a notable gap in existing studies. It conducts a thorough systematic review of a wide variety of surveys from multiple countries, examining survey and consent design characteristics to assess their influence on consent rates. Through statistical analysis, the study evaluates numerous factors such as survey response rate, topic, country, year, type, mode, age of target population, sponsor, and various aspects related to the consent request process. The findings suggest that higher consent rates are more likely in computer assisted telephone interview (CATI) compared to self-administered surveys, in surveys sponsored by governments, in panel surveys as opposed to cross-sectional ones, and where explicit consent is sought, especially when the consent request is made at the start of the survey. This research makes a significant contribution to academic literature by empirically demonstrating the effect of various survey and consent design features on consent behavior, providing essential insights for crafting surveys to maximize consent rates for data linkage. It closes an important research gap in survey methodology and the behavior on data linkage consent, offering a unique, extensive analysis across a broad range of international surveys.

Jel Codes: *C89, C42, C14*

Keywords: *Data Collection, Survey Methods, Semiparametric and Nonparametric Methods*

[1] Central Bank of the Republic of Turkey, Department of Statistics, turknur.hamsici@gmail.com, ORCID: 0009-0002-7211-7390

[2] Prof. Dr., Hacettepe University, Institute of Population Studies, Department of Social Research Methodology, aturkyil@hacettepe.edu.tr, ORCID: 0000-0002-2783-932X

Citation/Atıf: Brand, T. & Türkyılmaz, A. S. (2024). Investigating the Determinants that Influence Consent Behavior for Linking Survey Data with Administrative Records. *Fiscaoeconomia, 8*(2), 495-516. Doi: 10.25295/fsecon.1412153

Brand, T. & Türkyılmaz, A. S. (2024). Investigating the Determinants that Influence Consent Behavior for Linking Survey Data with Administrative Records. *Fiscaoeconomia, 8*(2), 495-516.
Doi: 10.25295/fsecon.1412153

Öz

Bu çalışmanın amacı, anket katılımcı verilerinin idari kayıtlar benzeri diğer verilerle birleştirilme aşamasında verdikleri onay davranışına etki eden tüm faktörleri araştırmaktır. Çalışma verisi, bu onayı verme oranı hesaplanmış ya da hesaplanabilen, uluslararası literatürde İngilizce olarak kaleme alınmış farklı alanlardaki anketlerin gözden geçirilmesi ve sistematik bir biçimde incelenmesi sonucu oluşturulmuştur. Diğer verilerle birleştirme onay verme oranı, o dönemin onay verenlerinin tüm anket katılımcılarına bölünmesiyle elde edilir. Bağımlı değişken onay verme oranı ve bağımsız değişkenler, anket ve onay verme süreci karakteristikleridir. Sistematik derlemenin sonrasında tüm değişkenlerin (anket cevaplama oranı, anket konusu, anket yılı, anket veri toplama metodu ve anket veri birleştirme onay isteme metodu gibi ilişkili degişkenler) değerlendirildiği istatistiksel modellerle onay verme davranışının belirleyicileri tartışılmıştır. Anketörle, bilgisayar destekli telefon anketlerinin, anketörün olmadığı anketlere göre; devlet sponsorlu anketlerin, özel anketlere göre; panel anketlerin, yatay kesit anketlere göre ve onay verme sorusunun basta sorulduğu anketlerin, diğer bölümlerde sorulanlara göre daha yüksek oranda verilerinin birleştirilmesine onay verdikleri bulunmuştur. Çalışmanın sonucu, diğer verilerle birleştirme konusunda yüksek oranlı onaya ulaşmanın anket katılımcılarının güven oluşturma süreçleri ile paralel davranış sergilediklerini düşündürmektedir. Araştırmanın bulgularının bundan sonraki onay verme süreci incelemelerine de bir perspektif kazandırabileceği görülmektedir.

Jel Kodları: *C89, C42, C14*
Anahtar Kelimeler: *Veri Toplama, Anket Metotları, Yarı Parametrik ve Parametrik Olmayan Metotlar*

Brand, T. & Türkyılmaz, A. S. (2024). Investigating the Determinants that Influence Consent Behavior for Linking Survey Data with Administrative Records. *Fiscaoeconomia, 8*(2), 495-516.
Doi: 10.25295/fsecon.1412153

1. Introduction

In the last century, the interest in accessing as much data as possible in every sector of life has become very popular and resulted in huge demand for data research (Kim & Rao, 2012; Fobia et al., 2019). Therefore, combining survey and administrative data records aims to enhance the amount of information available in each unit and support analysis for research purposes. Government departments and other organizations collect administrative data for registration, transaction, and record-keeping purposes, while survey data is based on a more targeted population and includes specific questions. If the administrative and survey data can be merged, the results are more beneficial. First of all, if the identifier information is correct, missing survey data may be completed. Secondly, utilizing administrative data with the intention of gathering full population records can lead to shortened sample questionnaires, thereby reducing time, cost, and respondent burden. Therefore, when feasible, data linkage is widely utilized.

However, the initial step to accomplish a linkage between survey and registered data is to obtain permission from participants to use their private and confidential information for any purposes. Taking ethical concerns into consideration, it is necessary to obtain the consent, i.e. approval, of the data owners (individuals or other organizations, such as establishments), who are participants of the survey, for confidentiality purposes. For the combination of survey and administrative data, obtaining consent is a preliminary step before requesting permission to link the data from survey participants. Such permission can result in either approval or disapproval. The proportion of respondents giving consent to link their data to all survey respondents in related period is called consent rate for data linkage. At this stage, it is very important to communicate the data linkage to the public to help them understand how important it is for public policy making.

Generally, accessing information from consenters is easier than from non-consenters, as their data is not linked. Normally, not all sampled participants provide consent. There are consenters and non-consenters. The survey data from consenting respondents is linked to the administrative data. Not receiving a positive answer to record linkage consent can create several problems in a survey once survey data is required to be linked to the administrative data. The first problem starts to arise when non-consenters have different characteristics from consenters, which can result in biased survey estimates (Kho et al., 2009; Sakshaug & Kreuter, 2012; Sakshaug et al., 2017). The second issue with non-consenters is that it diminishes the effective sample size because fewer units are connected to the administrative data than anticipated, as confirmed by the survey. As a result, the sampling variance increases, and the accuracy of the estimates decreases. If the consent rate can be predicted in advance by using a past survey on a similar topic, the initial sample size may be inflated accordingly. Increasing the sample size decreases sampling variance, thereby correcting for randomly distributed non-consent. However, it does not reduce any bias introduced by systematic non-consent. In this study, it is aimed to review surveys asking consent for data linkage in the international literature and to find out what has been influential for survey respondents to permit to link their survey data with other records available, in other words, what are the significant factors

of giving consent for linkage structured as survey design and consent characteristics variables in this research (section 3.4).

Although linking administrative and survey data has been theoretically simple, the practical implementation of finding all matching records by names, dates of birth, or other identifying information has proven to be exceedingly complex. Particularly when attempting to link millions of population records together. This study examined the identifiers of consent for linkage through a systematic review and analysis, as outlined in detail in the following section2. Following the Preferred Reporting Items for Systematic Reviews (PRISMA) guidelines (Moher et.al., 2010), first, a systematic literature search has been included reading through eligible manuscripts and at the same time this has been tool to collect survey and consent design data with calculated consent rate (section 3.3). During the second phase, the study compiled data from various surveys to conduct a regression analysis. The study aimed to explain consent behavior and sought to answer the following research questions:

- Which survey design characteristics significantly impact the rate of consent?
- Which characteristics of consenting to data linkage are significant for the consent rate?

This paper is structured as follows. The next section includes a literature review on consent rates for survey and administrative data linkage, and factors that are known to influence these rates. The second section explains the methodology including research strategy, inclusion exclusion criteria, survey selection, data set and preparation of variables for statistical analysis. In the third section, the results are evaluated. Afterward, all descriptive and regression analyses implemented are discussed. In the final section, results, limitations, and present implications for further work on consent rates are discussed.

2. Former Studies

Since the beginning of linking survey and administrative data, there have been ongoing discussions in the literature about who gives consent and who does not. Although there has been a great deal of research on consent behavior for data linkage over the past decade, there is still no consensus in the literature that explains the mechanisms of survey and consent design (Baghal et al., 2019; Carter et al., 2010; Sakshaug & Kreuter, 2014; Sala et al., 2012). Studies in the literature have varied in populations, modes, and consent methods, so the results of one study cannot be directly generalized to other studies (Kreuter, Sakshaug & Tourangeau, 2016).

So far, the placement and wording of the consent question have been the leading characteristics observed. Sakshaug & Vicari (2018) found that in a web survey of establishments, a consent question for linkage placed at the beginning of the survey received a 61.3% positive response. Also, in a CATI survey of the employed/unemployed, the consent rate is 95.6 percent when the question is asked at the beginning of the survey (Sakshaug, Tutz & Kreuter, 2013). On the other hand, placing consent questions somewhere in the context of computer assisted personal interview (CAPI) survey of households results in a 65 percent

consent rate (Sala, Knies & Burton, 2014). Although the placement of the consent question is effective, the mode of the survey is also influential.

Grove's Leverage Saliency Theory (2000) explains differences in survey participation by making the survey more appealing to respondents through methods such as incentives. Most of the information in the consent form could help participants make a decision. According to this, placing relevant linkage consent questions in a beneficial frame may make them more interesting for respondents (Sakshaug, et al., 2013). What is said to explain benefit framing and how the respondent interprets it as beneficial plays a key role. Is the reason for consenting to data linkage time-saving by answering fewer questions, a scientific purpose, linkage helps for an unknown answer, or something else? For busy respondents in a CATI employee survey, time gain information is effective in obtaining a higher consent rate than neutral questioning (Sakshaug et al., 2019). However, negative wording as “less useful” emphasizing the loss in a CATI survey of US registered voters resulted in a 66.8% consent rate (Kreuter, Sakshaug & Tourangeau, 2016). It all depends on the human factor, the respondents' perception of their role in life. Elevelt (2021) reviewed some of the variables structured under network meta-analysis and descriptive reviews. The survey sponsor and the wording are found to be influential in the consent rate. However, only experimental studies were included in this review, and some reviews are not strong enough because the research sample size is too small.

3. Methodology

3.1. Research Strategy

In this study, articles using survey data asking consent for linkage were screened on the Google Scholar, Web of Science, EBSCO host, JSTOR, ProQuest, Science Direct databases that were published in English, available in full text as a report or published paper between 1946 to 2023. In this context, to analyze the phenomenon of consent rate, the study had at least one survey data with a consent question for linkage, which has been asked to the sampled respondents and a consent rate has been calculated or possible to be calculated by using survey data. Additionally, key phrases such as “survey”, “authorization”, “permission”, “consent”, “linkage”, “combine”, “match” were employed. By examining the references of found research enabled to access other surveys with consent linkage question which is called snowballing research. Similarly, the following studies have been processed further in the same method. There have been cases where those studies lacked the necessary information regarding the survey and consent characteristics. This information was first searched on the internet, if it was not available then it was computed from the existing dataset. Further investigations on those surveys have led to access to reports, other papers, data sources, etc. to be able to fill out the corresponding missing information. Otherwise, it was requested from the authors of the papers and survey staff contacts in the relevant organization via email.

This review is conducted according to the Moher et al. (2009) guidelines on PRISMA steps indicated below in section 3.2. Following that statistical procedures are used to describe the results of a number of research studies. It is one way to summarize, integrate, analyze and

interpret selected sets of measures in various disciplines. It applies only to empirical studies; it cannot be used to summarize theoretical work. Second, it applies to research studies producing quantitative findings; reporting descriptive or inferential statistics with quantitative variables to summarize results. Third, being a technique for comparing same statistics of different studies, it is necessary that comparisons are meaningful, not apples and oranges. The results of each survey are calculated to obtain an overall estimate of the magnitude of effects and assure comparability. This is called effect size, a statistic like correlation, rate, etc. summarizing the construct researched and compared which is the consent rate for this study. While Hedges (1984) tried to structure the type of meta statistical analyses like significance tests, etc. on interpreting effect sizes, according to Glass (1976), answering new questions with secondary data using any statistical technique is so-called meta-analysis.

In the review process in each survey, first of all, the titles and abstracts of the articles were examined and titles were saved in an Excel file. Articles, being a resource to reach survey data with consent, relevant survey data were found in other online resources for the missing part. Independent variables of survey and consent characteristics were created after discussions of literature and recorded as well in the file. Finally, statistical analyses were used to investigate the consent identifiers in a general sense.

3.2. Inclusion and Exclusion Criteria

A prior eligibility criterion for all survey data found and collected was that the researchers be able to calculate consent rates or have enough information to calculate consent rates for data linkage. The article should be written in English with consent asked for data linkage.

Data naturally included different types of surveys such as longitudinal surveys, cross-sectional surveys, and experimental research. The basic criteria for the data were to obtain information over a period of time about the respondents who were asked the consent question for the first time in that particular survey. Therefore, for survey types where respondents were followed over time, such as longitudinal surveys, the first wave of the survey was used. However, if there was a newly recruited refreshment sample or a methodological change in the survey that was likely to affect consent behavior over the years, the first wave of that new period was also included in the dataset. The Panel Study of Income Dynamics (PSID) and the British Household Panel Study (BHPS) are some examples of such surveys.

Different groups of respondents may participate in each period due to the sampling design of cross-sectional surveys. Therefore, consent for data linkage was required from these participants, and consent rates were recorded in each period. Also, experimental research processed by using the survey data with a consent research question and variables involved in the data. For example, in Sakshaug et al.'s (2017) Legitimation of Inequality over the Life Span (LINOS) study, the entire employment survey sample was randomly assigned to interviewer-administered (face-to-face) or self-administered (mail/web) interviews that included consent questions linked to federal employment records. The purpose of this experimental study was to determine the differences in consent rates and consent bias among the survey modes tested for the entire sample when requesting consent for record linkage. Therefore, they were treated as separate data observations in the data.

Exclusion criteria were as follows. If a study simultaneously asked for survey participation and consent for linkage, it was excluded. In other words, if they were part of the same request, those studies were not involved. As it could be diagnosed by the method itself, in such surveys, there was not any possibility to calculate consent rates and evaluate consent behavior.

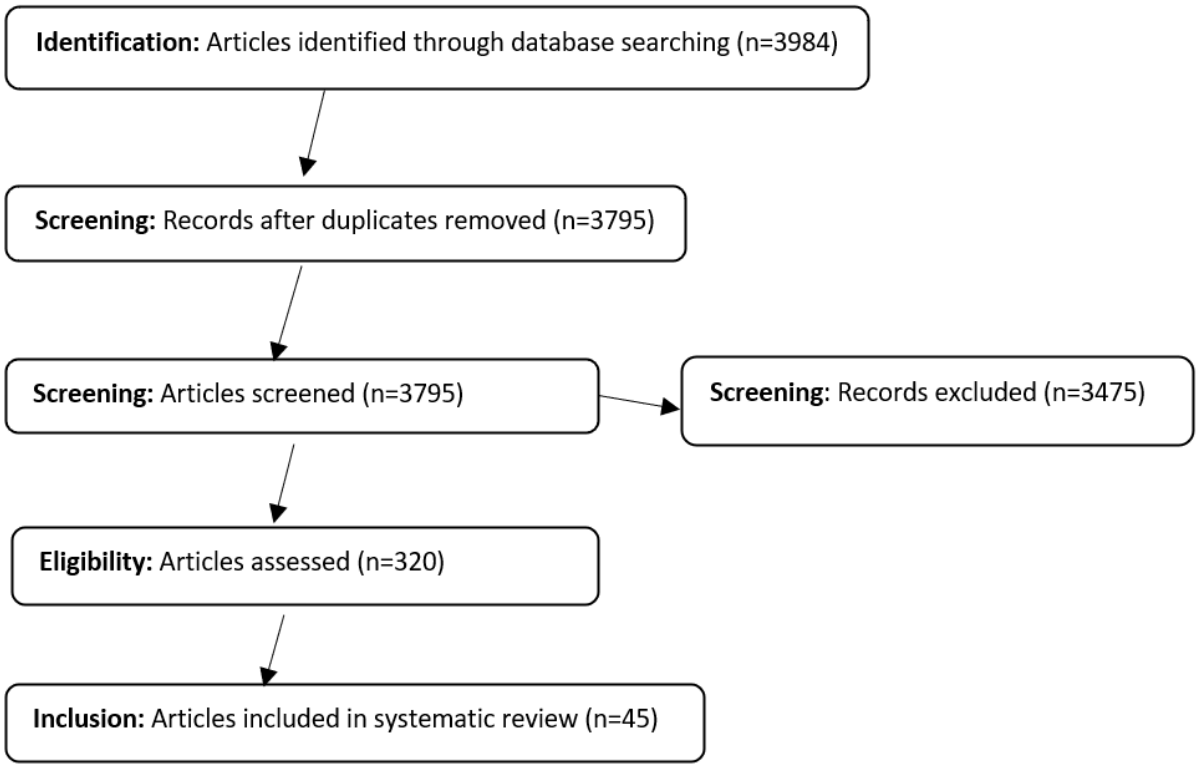
As expected, more than one article was examined using the same survey data. The strategy was to mention those articles with more information that were examined when creating variables. The others were used as a supplemental resource for missing information about the survey included. There have also been cases where the article was first studied as a working paper and later published in a journal. In this situation, published research was preferred only when there was enough of the required information. After all these steps, 45 articles were included in the systematic review.

3.3. Survey Selection

Surveys were selected according to the PRISMA checklist (Moher et al., 2010). This checklist is a specific procedure for organizing survey research and converting observations into a dataset. An initial list of 3984 references was produced by using the electronic search. Examination of the titles and abstracts, and further examination of the reference lists in the retrieved papers and other sources, led to a preliminary list of 45 potentially relevant articles in Germany, the United Kingdom, the United States, Canada, Australia, and New Zealand.

189 out of 3984 articles were eliminated because they were exact duplicates of the same article in different article references. After reviewing the remaining 3795 articles, several situations were identified. Linkage error has been studied when there are different matching procedures implemented when reaching out to larger administrative data files with more variables needed. There were some other consent forms, that is informed consent, asking for permission to participate in the survey. Physical measurement permissions for health research and health surveys had a different type of consent. As a result, 3475 articles were excluded and 320 were in scope. Of these 320 articles, 275 cases were removed because they were duplicates or asked both informed and linked consent questions at the same time in the survey. Through the systematic review of articles, papers, and reports in the literature, the information of 53 surveys were collected from 45 articles.

Figure 1: PRISMA Flowchart (Article Identification and Eligibility Assessment)



Using the data set construction strategy, these studies retained 45 articles and 158 survey data observations, in other words, these consent rates were extracted using the eligibility criteria described above. As expected, all of these surveys had different methodologies, depending on the goal of the survey research. Each of these studies found through the use of keywords was examined to determine whether it was a methodological report of a repeated cross-sectional or panel study, or a specific experimental project of a continuing or newly constructed survey project for this research. All studies reporting surveys with consent for linkage in all areas were included, including experimental research on these surveys. Although only English language resources were included in the dataset, German language resources were also used, when necessary, as a source of missing information needed to construct variables. Within the abstracts reviewed, if a study with consent to link was found, but not all information could be accessed, other resources, reports, or as a last step, authors were contacted to obtain the relevant information. Three authors were contacted. The government institution was consulted for missing data for one survey.

3.4. Dataset Preparation and Variables

This section discusses the variables that were created to be examined in statistical models to explore their effects on explaining consent rates. The variables are survey and consent design characteristics in categories. Survey response rate and consent rate are continuous variables. The categories of variables created are shown in Table 1.

Survey Topic: For each type of consent request, the consent rate for different survey topics shows variability across surveys. In most cases, health surveys have a higher rate of informed consent to participate in the survey (Voogt & Van Kempen, 2002), but according to an experiment conducted by Keusch et al. (2019), the topic has no effect on the relevant consent. In this research, surveys with topics of education, income, health, and others were categorized as a variable

Survey Country: Given the number of factors that are likely to be at play in countries, it would be very ambitious to claim that one country has more or less consent to data linkage than another. Through their socio-economic structures, cultural characteristics, and thus their laws, they develop different attitudes towards consent, privacy, and data sharing issues. Therefore, it is difficult to assess whether the variability in the laws and social and cultural conditions related to consent causes a major difference in the consent rates of Germany, the United States, Canada, Australia, New Zealand, and the United Kingdom in this study.

Survey Year: Compared to previous years, the linking of survey and administrative data has become more common, and more sophisticated procedures have been developed to obtain consent. Although respondents were more cooperative, it is unlikely that increased awareness of privacy and confidentiality issues will lead respondents to be more secure with their data, resulting in lower consent rates for data linkage. It is not easy to know whether surveys from older or later years have higher consent rates.

Survey Type: In cross-sectional surveys, consent for linkage is asked at each wave of the survey; in panel surveys, the sample constructed in the first wave of the survey responds to consent. As clarified in the consent question, they can change their decision at any time in future waves of the survey once they have been reminded of their consent to data linkage.

Panel respondents are expected to start with trust in the organization conducting the survey and build more trust over time (Sakshaug et al., 2012). However, respondents usually do not know or pay attention to whether it is a panel or cross-sectional survey until the second wave. That is, it is unclear at what stage and to what extent they will be informed about the survey type, which may determine their decision to consent to data linkage. Thus, the question is whether or not survey types differ in their consent rates for data linkage.

Survey Target Population Age: There is inconsistency across studies in how some survey characteristics affect consent behavior. The age of the participants is one of the characteristics that has been shown to be effective, possibly in both positive and negative directions. For example, some studies find that the older age group is more likely to consent to data linkage (Bryant et al., 2006; Dunn et al., 2004), while others find that the younger age group is more likely to consent to data linkage. (Hung et al., 2007; Yawn et al., 1998). In addition, some studies have found no effect of age on consent for linkage (Buckley et al., 2007; Harris et al., 2005).

Survey Mode: It is emphasized that it is necessary to explain the data linkage to survey respondents by giving as much information as possible with more clear clarifications (Thornby et al., 2017) which is guaranteed on interviewer administered surveys. There is an opportunity for participants in a survey with an interviewer to ask clarifying questions; having an

interviewer is a highly preferred option for these groups of respondents, as observed in an Understanding Society longitudinal survey. (Jäckle et.al, 2018). In a randomized mode study of the Legitimation of Inequality over the Life Span (LINOS) panel survey in Germany, the CAPI survey mode had about 40 percentage points higher linkage consent rates than the self-administered survey mode (Sakshaug et al., 2017).

In this study, to determine the effect of personal contact with interviewers on consent behavior, surveys were categorized as CAPI, CATI, self-administered, dual mode (CATI and CAPI or CAPI and Web), sequential mode. Sequential mode is the case where a mode type, such as self-administered, is first implied and then telephone mode is sequentially implied to the same sample after the previous mode failed to reach the respondent.

Survey Sponsor: The survey sponsor is usually identified to respondents during the initial contact when the purpose of the survey is explained and the consent forms request a response to the survey. In general, government-sponsored surveys may be perceived as compulsory to some degree, probably depending on the governmental system in different countries. On the other hand, non-governmental surveys tend to give respondents more flexibility in making their own decisions about whether or not to participate. Assuming that the government-sponsored survey happens to have a positive return on response (Heberlein et al., 1978), some research shows that government-sponsored surveys achieve higher response rates. (Linsky, 1975; Goyder, 1985; De Leeuw & Heer, 2002). It is believed that respondents may have stronger bonds of trust and respect for their governments, or feel a sense of obligation to participate, as well as for the consent data linkage.

Survey Consent Request Type: There are two approaches, either explicit (active) or implicit (passive), to obtaining survey consent from participants. In the explicit consent condition, subjects are asked to sign and return a consent form documenting their permission for their contact information to be transferred from administrative sources to third-party data collection agencies. In the implicit consent method, however, permission is granted if no action is taken. In other words, the signed opt-out form is not returned; the respondent has the responsibility to opt out at some point in the survey process. Consent for data linkage is the default option.

Linkage Request Placement: There is no definitive agreement on where to place the consent question in the survey, although most studies have placed the consent question closer to the end of the questionnaire. Linkage consent is interpreted as a sensitive topic, first interviewers build rapport and respondents need to trust before asking for consent to data linkage. (Sakshaug et al., 2013).

Linkage Request Records: In most surveys, the records being asked for consent linkage are on the same topic as the study. Although the need for linkage to administrative records arises mainly for employment and health surveys, there are also other records, such as education, income and tax credit records, that may be requested for linkage in the context of surveys on the same or different topics. Thus, it is questionable whether or not the consent rate would be higher if the record were about health or not.

Consent Wording: The wording of the consent request and the information provided to explain the consent to data linkage can play a significant role in respondents' attitudes. (Sakshaug et al., 2013). It is well known that a strong guarantee of data confidentiality leads to higher response rates. However, this effect can also be reversed (Singer et al., 1992, 1995). Therefore, it is discussed that framing the consent question neutrally versus negatively or positively will have different influences on respondents to give a positive return.

Survey Consent Signature and Identity Information: The decision to sign a linkage consent form can also be examined through theories of informed consent for survey participation, with the same concerns about privacy and confidentiality. If the survey is on a sensitive topic that involves income, finances, blood test questions, or something similar, the situation may become even more difficult as they are informed that their data can be much more easily linked to administrative data (Jenkins et al., 2006; Sala et al., 2010). If respondents behave in a similar way as they do while giving consent to participate in a survey with a signature, then requiring a signature for linkage consent may be disadvantageous compared to not requiring a signature. In addition, asking for identity information may make it more difficult to obtain a positive consent response. Referring to those, either signature or ID; at least one; both categorized as variable.

Survey Response Rate: Responding to a survey and allowing the linking of one's registered data are not necessarily similar issues when both are investigating their questions. Taking a role in a survey is much easier to accept than sharing relatively private registered information. Respondents may view participation in a survey as beneficial to scientific research to the extent that their confidential information is not shared through linkage. Also, some respondents are comfortable doing both. That is, consent to take a survey and consent to link data may not be relevant at all. Sometimes with lower response rates, respondents are likely to be more selective and cooperative. Because these respondents are a group that fully trusts and engages with the survey, and naturally a higher percentage of them are likely to agree to link.

Table 1: Descriptives of Constructed Variables (Factors) with Categories

Variables	Categories	Counts	Percent (%)
Survey Topic	1: Other (ref.)	17	13,3
	2: Economic	33	25,8
	3: Health	78	60,9
Survey Country	1: Australia, NZ, UK (ref.)	31	24,2
	2: US, Canada	75	58,6
	3: Germany	22	17,2
Survey Year	1: Year >=2010 (ref.)	54	42,2
	2: 2001 < Year <2009	44	34,4
	3: Year <=2000	30	23,4
Survey Type	1: Cross sectional (ref.)	79	61,7
	2: Panel, longitudinal, cohort	49	38,3
Survey Mode	1: Self-administered (ref.)	26	20,3
	2: CATI only	53	41,4
	3: Other (dual, sequential)	10	7,8
	4: CAPI only	39	30,5
Survey Target Population Age	1: All age groups (ref.)	44	34,4
	2: Age <18	46	35,9
	3: 18<= Age	38	29,7
Survey Sponsor	1: Nongovernment (ref.)	43	33,6
	2: Government	85	66,4
Survey Consent Request Type	1: Implicit/passive/opt-out (ref.)	20	15,6
	2: Explicit/active/opt-in	108	84,4
Linkage Request Placement	1: At the end	84	67,2
	2: At the middle & before consent	41	32,8
Linkage Request Records	1: Other (ref.)	31	24,2
	2: Economic	25	19,5
	3: Health	72	56,3
Consent Wording	1: Framing (ref.)	56	47,5
	2: Neutral	62	52,5
Consent Signature and Identity Info	1: Signature and idinfo (ref.)	5	4,1
	2: Signature or idinfo	49	40,2
	3: None required	68	55,7

3.5. Statistical Methods

In this research, the characteristics of surveys that can be influential in explaining the consent behavior for linking survey and administrative data were investigated. Based on previous survey research, the potential impact of various features of survey design and consent request methods were estimated. Those evaluations were based on data gathered through systematic reviews of prior survey studies. After examining literature including questionnaires, publications, and methodological documentation, it became apparent that certain survey features greatly impact obtaining consent approval. To analyze these features, the characteristics outlined in section 3.4 were utilized in generating categorical variables that were tested via statistical models, with the consent rate representing a dependent variable.

The consent rate was defined as the percentage of respondents who authorized the linkage of their survey responses to registered data over the ones who did not. During the construction of the models, the survey design and consent variables were analyzed and described as independent, exploratory variables. For each independent variable, categories were established based on the number of observations in each category and those demonstrating disparities in consent rates. The reference category was chosen as the one that is frequently compared in the literature for its response to consent for data linkage.

Initially, descriptive statistics, including frequencies, means, median, were evaluated for all variables. They were also checked for correlations with Kruskal Wallis test to describe their relationships with each other if there exist. Consent rates for each of these variables through binary regressions were estimated. Then multiple regression analysis with all variables was conducted. For final model, multicollinearity was assessed by computing a variance inflation factor (VIF) which measures how much the variance of a regression coefficient is inflated due to multicollinearity in the model. If VIFs are above 10 then that indicates multicollinearity (Dormann et al., 2013; Hair et al., 2014).

4. Results

Within the scope of the study, 128 calculated consent rates for data linkage on various topics with different design characteristics from 45 articles were systematically analyzed. The consent rate was 74.9% on the overall average, with a minimum consent rate of 15.4% and a maximum of 100%. The survey response rate was 64.2% on average, with a minimum response rate of 9.23% and a maximum of 96%.

Those variables have been studied by using the data collected in certain criteria, which were explained in section 3. Kruskal Wallis test proves that the variable categories are significantly different from each other in giving consent for linkage. Estimates, standard errors and VIFs of variables taking part in the regression model are presented in table 2. The closer to the smallest possible value of VIF, which is 1 meaning no correlation, the moderately the predictors are correlated.

The results on regressions examined suggest that six variables explained below -survey mode, target population age, linkage request placement, survey type, survey sponsor, response rate- were significantly associated with the consent rate. It was found that surveys, which were panel, interviewer-administered, government, targeting younger populations, and request for consent linkage placed at the beginning and before consent are positively related to receiving a higher consent rate.

4.1. Main Findings of Variables

The issues discussed in the studies that were included in the analyses using consent rate as a dependent and other variable created as independent variables. The amount and direction of effect of those variables in other words factors were explained below (Table 2).

Survey Topic, Country, Year, Type and Target Population Age: were some of the exploratory variables in the models that thought to be not influential. It is not expected that health surveys as a linkage request record will not have a higher rate of consent approval for linkage compared to the other records asked in the survey. That is, both the survey topic and the linkage request record are not statistically significant in explaining the data linkage consent rate. Within these, survey type and target population age have some significance in the regression model explaining consent rates. There are 49 longitudinal and 79 cross-sectional surveys in the data set. The average consent rate of longitudinal surveys is 71.9% and cross-sectional surveys are 77%. longitudinal surveys have a higher return of consent than cross-sectional surveys, likely due to the confidence maintained in the longer period. There are 29.7% of surveys specifically targeting 18 years and older; 35.5% of surveys targeting below 18 years; 34.4% surveys targeting all age groups. The younger population targeted surveys are significantly associated with an average increase of 15.8 in consent rates compared to the older population targeted ones, showing a higher consent linkage approval in younger populations than in older populations.

Survey Mode: Interviewer administered surveys - CATI only, CAPI only, dual &sequential had higher consent rates compared/reference to self-administered surveys. There was 39 CAPI, 53 CATI, 26 self-administered surveys, and 10 surveys on dual and sequential modes in the data with an average consent rate of CAPI, 62.9%; CATI, 86.2%; self-administered, 68% and others 77.5%. CATI surveys are significantly associated with an average increase of 18.2. In summary, CATI mode surveys have on average higher consent rates compared to CAPI, in comparison to self-administered interviews.

Survey Sponsor: There were 85 government-sponsored surveys with an average consent rate of 79.1% and 43 nongovernment surveys with an average consent rate of 67.2%. Government-sponsored surveys had on an average increase of 13.7 in consent rates compared to nongovernment surveys.

Survey Consent Request Type: There were 20 surveys with an implicit consent request type with an average 52.9% consent rate 108 surveys with an explicit consent request type with an average of 79.6% consent rate. Surveys with explicit consent request types were significantly associated with an average increase of 26.9 in consent rates compared to implicit ones.

Linkage Request Placement: There were 41 surveys with the placement of linkage request questions before consent questions with an average consent rate of 87.4%; 84 surveys with the placement of linkage request questions at the end with an average consent rate of 68.6%. The regression model showed that surveys, where the linkage request placement was at the beginning or before the consent question had significantly an average increase of 10 in consent rates compared to surveys with linkage request placement at the end.

Survey Response Rate: The survey response rate and consent rate have a Spearman correlation coefficient of -0.15, indicating that there is almost no association between those two variables. Pearson correlation was checked for the two numerical variables, survey response rate and consent rate. As seen, Table 2 shows that the regression coefficient of

response rate is -0.32 with a standard error of 0.13, which also means that there is a negative association between response and consent rates.

Table 2: Multiple Regression Analysis Results of Consent Behavior Variables

Variables	Categories	Estimates	Std.Err	VIF
Survey Topic	1: Other (ref.)			
	2: Economic	-3,94	7,75	1,73
	3: Health	2,68		
Survey Country	1: Australia, NZ, UK (ref.)			
	2: US, Canada	1,39	6,52	2,39
	3: Germany	-11,67	11,19	
Survey Year	1: Year >=2010 (ref.)			
	2: 2001 < Year <2009	0,59	4,48	1,36
	3: Year <=2000	-1,5	5,26	
Survey Type	1: Cross sectional (ref.)			
	2: Panel, longitudinal, cohort	15,28**	4,59	1,61
Survey Mode	1: Self-administered (ref.)			
	2: CATI only	18,21**	6	1,45
	3: Other (dual, sequential)	1,59	7,62	
	4: CAPI only	1,07	5,87	
Survey Target Population Age	1: All age groups (ref.)			
	2: Age <18	15,78**	5,25	1,55
	3: 18<= Age	0,12	5,23	
Survey Sponsor	1: Nongovernment (ref.)			
	2: Government	13,68**	4,69	1,6
Survey Response Rate		-0,32*	0,13	
Survey Consent Request Type	1: Implicit (ref.)			
	2: Explicit	26,93***	6,93	1,82
Linkage Request Placement	1: At the end			
	2: At the middle & before consent	9,95*	3,91	1,31
Linkage Request Records	1: Other (ref.)			
	2: Economic	4,83	8,44	2,03
	3: Health	0,02	6,03	
Consent Wording	1: Framing (ref.)			
	2: Neutral	8,02	5,22	1,96
Consent Signature and IdInfo	1: Signature and Idinfo (ref.)			
	2: Signature or Idinfo	-12,43	9,38	1,64
	3: None required	-6,23	9,74	

Notes: *p<0,01 **p<0,001 ***p<0,0001

5. Discussion and Conclusion

In this study, a systematic review was performed to access survey literature with consent rate for linkage of survey and administrative data. The literature with this type of consent were

aggregated and summarized and several aspects of the consent question that affect consent rates were discovered.

Based on all of this information, four factors -survey mode, survey sponsor, response rates, and linkage request placement- were supported by the results of previous research. In addition, explicit versus implicit consent and written versus oral consent forms were different than previous. A higher consent rate for explicit consent requests, or no differences between oral and written consent forms, is likely related to some other characteristic dimensions that can be explored in further research.

Results on survey mode show that respondents still prefer human interaction, although new technologies such as online web surveys limit this to some extent. Even though it is not face to face, having a possibility for communication with interviewers about survey study via telephone (CATI) increases the chance to receive higher consent approvals. Interviewers can help respondents overcome their lack of knowledge and provide them with the descriptions they need to make an informed decision, which can ultimately lead to higher consent rates for data linkage. In addition to the security concerns in online platforms, the necessity to do their reading and understanding seems to be a task to do for the respondents which they are not always willing to do. Face-to-face contact with respondents works well to get consent for linkage, similar to getting a response (Jenkins et al., 2006).

Results on sponsors, linkage request placement, consent request type and survey type prove once again the importance of trust for the participants. As known, there can be a variety of reasons for taking part as a respondent, wherein on some occasions, government sponsored surveys with a sense of obligation to participate can be a stress factor for respondents or anything else (Linsky, 1975; Goyder, 1985; De Leeuw & Heer, 2002). Clearly indicating consent question at the beginning, maintaining a longer correspondence of more than one period as in panel surveys and government being the main organization to trust likely leads to higher consent approval for linkage. Asking for consent as early as possible for panel surveys like HRS is observed to result in higher rate of positive consent approval (Sakshaug, 2021). Longitudinal surveys have a higher consent rate than cross-sectional surveys. If the survey organization in panel surveys -by giving detailed information regarding consent- builds a rapport with respondents and also by mentioning the period of the survey, they already create some kind of trust. In addition to that, explicitly asking for consent rather than implicit also supports trust mechanism. Therefore, as further research, that would be interesting to look into interactions between some trust related variables once enough sample size is reached.

Although the literature has been trying to explain the consent issue in a similar framework to the non-response theory, the behavioral characteristics of survey participation and consent for linkage are not necessarily the same. The contribution of factors such as effective interviewers, advance letters, and refusal conversion efforts to unit response and willingness to consent to administrative record linkage is predicted to be positively related (Groves et al., 2004). As opposed to general belief, this research shows that lower response rate has a higher consent for linkage approval rate than surveys with a higher response rate. Respondents can be “willing” to answer the survey questions, but they are not fine with sharing their private information by approving the consent for linkage. Thus, even though the response rates are

high in such cases, consent rates can be less. Another argument would be that in a survey with a lower response rate, sample respondents are already resistant at the response-giving phase, the remaining who respond have a higher consent.

In summary, the results of our analysis strengthen the importance of trust built with the survey participants in every field. Trusting the survey, which they participate, opens the doors for survey researchers for correspondence and commitment to a following step as consent linkage which is their confidential information at the end. Thus, any survey study by using all possible resources should aim building trust with respondents at first stage to maintain positive returns for higher consent rates.

Typically, results are subject to some limitations. First, the research only in the language of English is considered. There is a possibility to have surveys with consent asked in other languages speaking countries as well. However, the consent request, which is the output of detailed research and strategies, is relatively more commonly applied in English-speaking countries. Thus, this is not a drawback. However, this study, being a systematic review with extra statistical analysis, is interpreted within the frame of mentioned studies. It is also the largest of its kind in the systematic review of consent rates.

Despite these limitations, results are valuable to researchers investigating the consent behavior for at least two reasons: 1) they are the first to offer a systematic review of this large sample on consent rates relying on the survey literature; 2) they identify some of the key factors related to building trust with survey organization that are associated with higher consent rates. Therefore, results can provide some useful insights for researchers when designing their survey studies.

References

Angus, V. C., Entwistle, V. A., Emslie, M. J., Walker, K. A. & Andrew, J. E. (2003). The Requirement for Prior Consent to Participate on Survey Response Rates: A Population-Based Survey in Grampian. *BMC Health Services Research*, *3*, 21. http://www.biomedcentral.com/1472-6963/3/21

Baghal, T. A., Knies, G. & Burton, J. (2014). Linking Administrative Records to Surveys: Differences in the Correlates to Consent Decisions. *Understanding Society Technical Report,* No. 2014-09. Institute for Social and Economic Research, Essex, UK.

Baghal T. A., Sloan, L., Jessop, C., Williams, M. L. & Burnap, P. (2019). Linking Twitter and Survey Data: The Impact of Survey Mode and Demographics on Consent Rates Across Three UK Studies. *Social Science Computer Review, XX*(X), 1-16.

Bates, N. (2005). Development and Testing of Informed Consent Questions to Link Survey Data with Administrative Records. *Proceedings of the Survey Research Methods Section of the American Statistical Association* (3786-3793).

Buckley, B., Murphy, A. W., Byrne, M. & Glynn, L. (2007). Selection Bias Resulting from the Requirement for Prior Consent in Observational Research: Community Cohort of People with Ischemic Heart Disease. *Heart, 93*, 1116-1120.

Carter, K., Shaw, C., Hayward, M. & Blakely, T. (2010). Understanding the Determinants of Consent for Linkage of Administrative Health Data with a Longitudinal Survey. *New Zealand Journal of Social Sciences, 5*(2), 53-60.

Cehovin, G., Bosnjak, M. & Manfreda, K. L. (2018). Meta Analyses in Survey Methodology: A Systematic Review. *Public Opinion Quarterly*, *82*(4), 641-660.

Das, M. & Couper, M. P. (2004). Optimizing Opt-Out Consent for Record Linkage. *Journal of Official Statistics*, *30*(3), 479-497.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., Garcia Marquez, J. R., Gruber, B., Lafourcade, B., Leitao, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D. & Lautenbach, S. (2013). Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance. *Ecograpy*, *36*(1), 27-46.

De Leeuw, E. & Heer, W. D. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. R. Groves, D. Dillman, J. Eltinge & R. J. A. Little (Ed.), *Survey Nonresponse*. New York: Wiley.

Dunn, K. M., Jordan, K., Lacey, R. J., Shapley, M. & Jinks, C. (2004). Patterns of Consent in Epidemiological Research: Evidence from Over 25000 Respondents. *The Practice of Epidemiology, 159*(11), 1087-1094.

Elevelt, A. (2021). *Consent to Data Linkage in Surveys: A Descriptive Review and Meta-Analysis*. Dissertation on Smart (Phone) Surveys.

Fobia, A. C., Holzberg, J., Eggleston, C., Childs, J. H., Marlar, J. & Morales, G. (2019). Attitudes Towards Data Linkage for Evidence-Based Policymaking. *Public Opinion Quarterly, 83*(Special Issue), 264-279.

Fulton, J. A. (2012). *Respondent Consent to Use Administrative Data*. Unpublished Ph.D. Dissertation, University of Maryland. http://drum.lib.umd.edu/handle/1903/13601

Glass, G. V. (1976). Primary, Secondary and Meta-Analysis of Research. *Educational Researcher, 5*(10), 3-8.

Gessendorfer, J., Beste, J., Drechsler, J. & Sakshaug, J.W. (2018). Statistical Matching as a Supplement to Record Linkage: A Valuable Method to Tackle Nonconsent Bias. *Journal of Official Statistics, 34*(4), 909-933.

Goyder, J. (1985). Face-to-Face Interviews and Mailed Questionnaires: The Net Difference in Response Rate. *Public Opinion Quarterly, 49*(2), 234-252.

Groves, R. M., Cialdini, R. B. & Couper, M. P. (1992). Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly, 56*, 475-495.

Groves, R. M. & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly, 72*(2), 167-189.

Groves, R. M., Presser, S. & Dipko, S. (2004). The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly, 68*(1), 2-31.

Groves, R. M., Singer, E. & Corning, A. (2000). Leverage-Saliency Theory of Survey Participation: Description and an Illustration. *Public Opinion Quarterly, 64*(3), 299-308.

Hair, J. F. (2014). *Multivariate Data Analysis* (Pearson New International Edition, 7th Ed.). Upper Saddle River, NJ, Prentice Hall.

Haraldsen, G., Jones, J., Giesen, D. & Zhang, L. C. (2013). Understanding and Coping with Response Burden. G. Snijkers, G. Haraldsen, J. Jones & D. K. Willimack (Ed.), *Designing and Conducting Business Surveys* (219-252). Hoboken, NJ: John Wiley and Sons.

Harris, T., Cook, D. G., Victor, C., Beighton, C., Dewilde, S. & Carey, I. (2005). Linking Questionnaires to Primary Care Records: Factors Affecting Consent in Older People. *Journal of Epidemiological Community Health, 59*(4), 336-338.

Heberlein, T. A. & Baumgartner, R. (1978). Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature. *American Sociological Review, 43*, 447-462.

Huang, N., Shih, S.-F., Chang, H.-Y. & Chou, Y.-J. (2007). Record Linkage Research and Informed Consent: Who Consents?. *BMC Health Services Research, 7*(18).

Jäckle, A., Beninger, K., Burton, J. & Couper, M. P. (2021). Understanding Data Linkage Consent in Longitudinal Surveys. P. Lynn (Ed.), *Advances in Longitudinal Survey Methodology*. Wiley Series in Probability and Statistics.

Jäckle, A., Sala, E., Jenkins, S. P. & Lynn, P. (2004). *Validation of Survey Data on Income and Employment: The ISMIE Experience*. ISER Working Paper, No. 2004-14. Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-14.pdf

Jenkins, S. P., Cappellari, L., Lynn, P., Jäckle, A. & Sala, E. (2006). Patterns of Consent: Evidence from a General Household Survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 169*, 701-722.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P. & Kreuter, F. (2019). Willingness To Participate in Passive Mobile Data Collection. *Public Opinion Quarterly, 83*, 210-235.

Kho, M. E. (2009). Written Informed Consent and Selection Bias in Observational Studies Using Medical Records: Systematic Review. *BMJ Research*. https://doi.org/10.1136/bmj.b866

Kim, J. K. & Rao, J. N. K. (2012). Combining Data from Two Independent Surveys: A Model-Assisted Approach. *Biometrika, 99*(1), 85-100.

Kreuter, F., Muller, G. & Trappmann, M. (2010). Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data. *Public Opinion Quarterly, 74*, 880-906.

Kreuter, F., Sakshaug, J. W. & Tourangeau, R. (2016). The Framing of the Record Linkage Consent Question. *International Journal of Public Opinion Research, 28*(1), 142-152.

Linsky, A. S. (1975). Stimulating Responses to Mailed Questionnaires: A Review. *Public Opinion Quarterly, 39*, 82-101.

Lugtig, P. (2014). Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers. *Sociological Methods & Research, 43*, 699-723.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. (2010). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The Prisma Statement. *International Journal of Surgery, 8*(5), 336-341.

Olson, K. M. (2006). Survey Participation, Nonresponse Bias, Measurement Error Bias, and Total Bias. *Public Opinion Quarterly, 70*, 737-758.

Pascale, J. (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). *US Census Bureau Working Paper, 2011*(3).

Sakshaug, J. W. & Antoni, M. (2019). Evaluating the Utility of Indirectly Linked Federal Administrative Records for Nonresponse Bias Adjustment. *Journal of Survey Statistics and Methodology, 7*(2), 227-249.

Sakshaug, J. W., Couper, M. P., Ofstedal, B. & Weir, D. R. (2012). Linking Survey and Administrative Records Mechanisms of Consent. *Sociological Methods and Research, 41*, 535-569.

Sakshaug, J. W. & Huber, M. (2016). An Evaluation of Panel Nonresponse and Linkage Consent Bias in a Survey of Employees in Germany. *Journal of Survey Statistics and Methodology, 4*, 71-93.

Sakshaug, J. W., Hülle, S., Schmucker, H. & Liebig, S. (2017). Exploring the Effects of Interviewer- and Self-Administered Survey Modes on Record Linkage Consent Rates and Bias. *Survey Research Methods, 11*(2), 171-188.

Sakshaug, J, W., Schmucker, A., Kreuter, F., Couper, M. P. & Singer, E. (2019). The Effect of Framing and Placement on Linkage Consent. *Public Opinion Quarterly, 83*(Special Issue), 289-308.

Sakshaug, J. W. & Vicari, B. J. (2018). Obtaining Record Linkage Consent from Establishments: The Impact of Question Placement on Consent Rates and Bias. *Journal of Survey Statistics and Methodology, 6*, 46-71.

Sakshaug, J. W. & Kreuter, F. (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, J. W. & Kreuter, F. (2012). Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data. *Survey Research Methods, 6*, 113-122.

Sakshaug, J. W., Schmucker, A., Kreuter, F., Couper, M. P. & Singer, E. (2016). Evaluating Active (Opt-In) and Passive (Opt-Out) Consent Bias in the Transfer of Federal Contact Data to a Third-Party Survey Agency. *Journal of Survey Statistics and Methodology, 4*, 382-416.

Sakshaug, J. W., Tutz, V. & Kreuter, F. (2013). Placement, Wording and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data. *Survey Research Methods, 7*, 133-144.

Sakshaug, J. W., Yan, T. & Tourangeau, R. (2010b). Nonresponse Error, Measurement Error, and Mode of Data Collection Method: Tradeoffs in a Multi-Mode Survey of Sensitive and Non-Sensitive Items. *Public Opinion Quarterly, 74*, 907-933.

Sala, E., Burton, J. & Knies, G. (2010). Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview and Interviewer Characteristics. Institute for Social and Economic Research, 2010-28.

Sala, E., Knies, G. & Burton, J. (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.

Sala, E., Burton, J. & Knies, G. (2012). Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview, and Interviewer Characteristics. *Sociological Methods & Research, 41*, 414-439.

Shahi, R. & Vousden, C. (2005). Bias from Requiring Explicit Consent from All Participants in Observational Research: Prospective, Population-Based Study. *BMJ*. Doi: 10.1136/bmj.38624.397569.68

Singer, E. (1978). Informed Consent: Consequences for Response Rate and Response Quality in Social Surveys. *American Sociological Review, 43*(2), 144-162.

Singer, E. (2003). Exploring the Meaning of Consent: Participation in Research and Beliefs About Risks and Benefits. *Journal of Official Statistics, 19*, 273-285.

Thornby, M., Calderwood, L., Kotecha, M., Beninger, K. & Gaia, A. (2017). Collecting Multiple Data Linkage Consents in a Mixed Mode Survey: Evidence and Lessons Learnt from Next Steps. Center for Longitudinal Studies Working Paper, 2017/13. doi:10.13094/SMIF-2018-00004

Tourangeau, R., Rips, L. J. & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge and New York: Cambridge University Press.

Webster, C. (1996). Hispanic and Anglo Interviewer and Respondent Ethnicity and Gender: The Impact on Survey Response Quality. *Journal of Marketing Research, 33*(1), 62-72.

Willimack, D. K. & Snijkers, G. (2013). The Business Context and Its Implications for the Survey Response Process. G. Snijkers, G. Haraldsen, J. Jones & D.K. Willimack (Ed.), *Designing and Conducting Business Surveys* (39-82). Hoboken, NJ: John Wiley and Sons.

Yawn, B. P., Yawn, R. A., Geier, G. R., Xia, Z. & Jacobsen, S. J. (1998). The Impact of Requiring Patient Authorization for Use of Data in Medical Records Research. *Journal of Family Practice, 47*, 361-365.

Ethical Approval: The authors declare that ethical rules are followed in all preparation processes of this study. In the case of a contrary situation, Fiscaoeconomia has no responsibility, and all responsibility belongs to the study's authors.
Author Contributions: author contributions are below;
Introduction: 1. Author %70, 2. Author %30
Literature: 1. Author %70, 2. Author %30
Methodology: 1. Author %70, 2. Author %30
Conclusion: 1. Author %70, 2. Author %30