RESEARCH ETHICS IN REGISTER-BASED POPULATION STUDIES

How do register researchers approach the ethics of working with human data?

Elisabeth Roxanne Spruijt (S2043556)
erspruijt@outlook.com

Supervisor: dr. A. Bailey (RuG)

Master Population Studies, Demography Department,
Faculty of Spatial Sciences, University of Groningen

Groningen, August 2017

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The scope of the field of population studies has broadened: there is an ever-increasing amount of micro-data available from registers to study population dynamics. Although register-based population studies tend to analyse de-identified private information, these studies are nonetheless accompanied by ethical concerns. Codes of conduct, meant to guide researchers making decisions in ethically ambiguous situations, do not include notions on this type of research, and might therefore be unfit to apply. The objectives of this study are twofold: first, to gain insight into the normative framework set out for academic researchers working in the Dutch context, and second, to explore how register researchers working in the Netherlands approach the research ethics of working with human data. A qualitative mixed method approach was employed, in which the findings of a content analysis on eight international, national, and institutional academic research ethics guidelines, and a thematic analysis of eleven semi-structured in-depth interviews with register researchers were combined to reach the objectives. The findings indicate that the norms set out in academic guidelines generally provide a suitable framework to understand researchers' perceptions and attitudes towards research ethics, but also reveal how researchers practice research ethics through additional strategies. These findings can be utilized when making policy about research ethics in register research. The developed inductive conceptual model forms a solid base for future research, as it provides valuable insights on the practices register researchers working in the Netherlands apply to protect the interests of those included in register datasets.

Keywords: population, register, research ethics, guideline, privacy, competence, Netherlands, mixed method

Chapter 1: Introduction

Population Studies is an interdisciplinary research field, mostly grounded in demography, but complemented with knowledge from for instance geography, sociology, psychology, economics, anthropology, and epidemiology. Population researchers study what might be peoples' most intimate moments; birth, romantic relationships, mobility, health, education, reproduction, work, and death. They do this by employing quantitative, qualitative, and mixed approaches, and using different data sources (primary or secondary collected data on local, national, or international level). While some of this research can be considered to be strictly fundamental, the larger share of studies serves as a basis for labour market-, housing-, health-, and education- policies on local and national level.

1.1 Problem statement

Two types of research are most common in population studies; human subject research, and non-human subject research. Human subject research is a systematic investigation, involving human subjects, designed to develop or contribute to generalizable knowledge (United States, 2016). It involves obtaining data through intervention or interaction with the individual and/or obtaining private, identifiable information of an individual. Whether a person is a human subject thus depends on the *type* of data that is collected in the research, and in what *ways* this is done. An intervention includes physical procedures of data collection and/or manipulations of the subject, or the subject's environment, performed for research purposes. With interaction, the data are gathered through communication or interpersonal contact between the researcher and the subject. Private information is "information about behavio[u]r that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place" or information which the individual provided for specific purposes, and "which the individual can reasonably expect will not be made public" (United States, 2016:1). Analysing private information (collected through intervention/interaction) is only considered human subject research when this information is individually identifiable, that is, it can be used to identify individual persons (United States, 2016).

Following the previous definitions, *non*-human subject research is research where humans are *not the subject* of the study (see also: The University of Utah, n.d.). The study's data are for example not collected through direct contact with the individual (intervention/interaction), or there is no identifiable, private information obtained. An example is register research, where population researchers work on municipal personal records data combined with other registers; large datasets that were obtained from for example municipalities or national government. The data used in these types of studies were not primarily collected for research, but for (semi-)government agencies to be able to fulfil their tasks (Rijksoverheid, n.d.). In the Netherlands, such microdata are for example collected by municipalities, the tax authorities, social security authorities, and education authorities. Under specific conditions, these data are shared with population researchers (Centraal Bureau voor de Statistiek, n.d.-a). The data analysed in this type of research tends to be aggregated or at least deidentified information on *human individuals*, which makes the term non-human subject research slightly confusing. Indeed, there can be no doubt that these data still relate to humans, albeit in deidentified form, which is why some researchers went on to define these studies as human non-subject research (Brothers & Wright Clayton, 2010). Defining it as such makes it evident that this type of research, alike traditional human subject research, comes with similar ethical concerns (Brothers & Wright Clayton, 2010). In this study, we will employ the terms human non-subject, and human non-subject research as (research on) humans, whose data are either collected without intervention or interaction with a researcher, or whose data are no longer identifiable or private.

1.2 Objective of the study

Despite the concerns related to human non-subject research, demographic literature has never given much attention to research ethics, as opposed to literature from sociology (Hedgecoe, 2008; Richardson & McMullan, 2007), geography (Dowling, 2009; Valentine, 2003, 2005) and psychology (Drotar, 2011; Koller, 2008). This absence of literature on research ethics from population researchers' perspective is problematic, especially since their work covers both human subject and human non-subject research and a combination thereof. As a result, population research can generate situations where it is unclear which set of research ethics to apply, and/or it can lead to instances where existing research ethics guidelines seem unfit.

This is also the case with register-based population studies, which on the one hand does not involve human subjects, but on the other, analyses with very personal information. There does not seem to be academic literature that can clarify how population researchers deal with this ambiguity in practice. However, codes of research conduct might be able to give some direction, as they aim to set a normative framework in which researchers make decisions about research ethics (Resnik, Rasmussen, & Kissling, 2015).

The first objective of this study is to get insight into the guidelines that were developed for academic population research in the Dutch context. Under this objective, the following research question was raised:

1. What normative frameworks are underlying the academic codes of research conduct?

The second objective is to explore how register researchers working in the Netherlands approach the research ethics of working with human data. Under this objective, three research questions will be studied:

1. What meanings do register researchers attach to human data research ethics practices as described in academic ethics guidelines, and what do they perceive to be the role of these practices in register research?

2. What are register researchers' perceptions and attitudes towards the ethical practices they apply in their work?
3. What role does researchers' competence play in forming perceptions and attitudes about research ethics practices?

This study aims to answer these questions, and to thereby add notions from the field of population studies to the existing body of literature on human non-subject research ethics.

1.3 Background

Population research involves different phases (e.g. literature study, data collection, presenting and publishing results) (Hennink, Hutter, & Bailey, 2011), and within and by transitioning between these phases, researchers make decisions about ethical issues (Guillemin & Gillam, 2004). Ethical issues are situations where the outcome of certain behavioural choices can harm or benefit others (Velasquez & Rostankowski, 1985). To deal with these situations, researchers abide by (inter)national laws, and norms established by the scientific community, which are often represented in Codes of ethical conduct. Codes of conduct aim to guide researchers when making decisions in ethically ambiguous situations (Resnik, Rasmussen, & Kissling, 2015) by distinguishing desirable and undesirable practices within the process of research (Koninklijke Nederlandse Akademie van Wetenschappen, 2012; Vereniging van Universiteiten, 2014). These codes exist on local, national, and international level, and are specified per research discipline, or applicable in all fields. The field of Population Studies does not have its own code, thus researchers working in this field consult either general guidelines, or specified codes from closely related fields such as psychology, anthropology, and sociology.

The Nuremberg Code (United States, 1949) is the result of what is arguably considered the first collective effort to establish a normative framework for research on human data. It was the result of the Nuremberg trials, where doctors who worked in Nazi concentration camps during the Second World War were tried for unethical and inhumane treatment of their patients (Shuster, 1997). The events that happened in those facilities demonstrated the importance of sufficient protection of human subjects in research. The judges of the trial came up with 10 research principles, focussing primarily on the rights of human subjects before, during and after participating in (medical) experiments, but suitable to apply to all kinds of research involving human subjects. Later, the influential Belmont Report was published (United States, 1978), aiming to provide an analytical framework that helps scientists resolve ethical problems that might arise from research involving human subjects. Previous codes of conduct were considered to be sometimes inadequate to cover complex situations, conflicting with each other, or hard to interpret (United States, 1978:1). Since then, numerous guidelines have been designed to guide researchers working on international, national, and institutional levels.

Various studies have been conducted to analyse the content of both business/organizational codes of conduct (e.g. Erwin, 2011; O'Dwyer & Madden, 2006) and international, national and institutional guidelines focussing more specifically on research conduct (Komić, Marušić, & Marušić, 2015; Resnik, Neal, Raymond, & Kissling, 2015; Resnik, Rasmussen, & Kissling, 2015). The analyses of codes of research conduct tend to yield similar results; they include notions on responsible research conduct, questionable research practices and research misconduct across different themes. Pimple (2002) grouped the different elements of research ethics into six domains; scientific integrity, collegiality, protection of human subjects, animal welfare, institutional integrity, and social responsibility. Clearly, not all domains are relevant for all types and disciplines of research. For example: while animal welfare might be relevant for pharmacists conducting animal experiments in the lab, and protection of human subjects might be relevant for those working in the humanities, this might not be true the other way around.

Human subject research tends to be a well-covered topic in codes of conduct. The guidelines are meant to protect the human subjects' rights, for example with regard to distribution of risks and benefits, and privacy protection. It is recommended that researchers planning to conduct human subject research present their proposals to an institutional review board (that gives non-binding guidance) before they start their study (Dolgin, 2014). In practice, most researchers cannot obtain funding or conduct a study without approval from their institutional review board. Human non-subject research, on the other hand, is a grey area with regard to regulations. Although providers of data tend to be strict in their assessment of requests, and institutions often still require approval from their review boards, academic codes of research conduct do not mention guidelines for human *non*-subject research.

This can be considered to be problematic, because these studies still deal with human data, and very personal information. Arguably, the risks of harm for human non-subjects whose deidentified data are included in research are less than they would be for traditional human subjects, but therefore not less significant. Non-subjects, and their data, should be protected as well, for example against the risk of re-identification, and the risk that their data are being used for other purposes than for which they were collected. While some (Rothstein, 2010) state that the current guidelines are insufficient in protecting non-subjects privacy, others argue that the fine line of de-identification and reidentification itself is sufficient protection for subjects (Brothers & Wright Clayton, 2010). The idea is as follows: human non-subject research falls under the category non-human subject research, which does not require permission from an institutional review board. In order for it to stay in this category, however, researchers must make sure that the data are properly deidentified, by processing and disseminating it in such a way that it remains non-human subject research (Brothers & Wright Clayton, 2010). Thus, with de-identification as dividing line between the two types of research, established codes of conduct should be guiding for human non-subject researchers as well.

But is this truly the case? Some have said that the codes of conduct used in social studies are too constraining, as they are developed for ‘strict’ biomedical sciences, while others feel they are not constraining enough, or just generally unsuitable to apply in social sciences (Israel, 2015). Fact is that some of the available guidelines are dated (the Nuremberg code stems from the 1940s, while the influential Belmont Report was written in the 1970s), and some authors have questioned the possibilities to compare these historical documents to present norms (Stark, 2007). Moreover, not only norms are subjected to change; the scope of the population researchers’ studies has changed as well as broadened.

There is an increasing amount of data available, and with new technologies to match and combine data sources being used more often, some even say data is so valuable, it is the new currency (Bakker, 2009; Sociaal-Wetenschappelijke Raad, 2006; Willenborg & Heerschap, 2010). These developments come with their own ethical concerns, for example with regard to the possibilities to analyse personal data that are seemingly freely available online. A Harvard study on relationships, culture and demographic characteristics of first-year students in the late ’00 demonstrated that the line between public and private information is not always clear cut (Israel, 2015). The researchers justified gathering data from students’ online social network profiles, by stating that data were not collected through interviews or other direct contact, and that the information was never made public. However, many were critical to this approach, as the information collected in the study was in most cases only available to closely related others in the social network, instead of publicly available (Israel, 2015).

1.4 Overview of the thesis

The previous example illustrates the ethical difficulties that come together with human non-subject research, and how current regulations might not be as guiding as they -perhaps- should be. This current study explores how population researchers approach these situations in practice. The conceptual framework used in this study (which is described in Chapter 2), was derived from a content analysis of eight codes of academic research conduct, while researchers’ perceptions and attitudes were studied through a thematic analysis of semi-structured in-depth interview data. An overview of the study design, and the methodological details of the data collection and analysis can be found in Chapter 3. The findings of this study are reported in Chapter 4. First, the meanings register researchers attach to human data research ethics practices as described in academic guidelines, and the role these practices play in their work are explored (4.1). Then, register researchers’ perceptions and attitudes towards the research ethics practices they apply in their work are discussed (4.2). Finally, the role that researchers’ competence plays, in the process of forming perceptions about ethics practices, is covered (4.3). The findings are discussed in Chapter 5.

CHAPTER 2: THEORETICAL FRAMEWORK

Register research can be considered a grey area when it comes to research ethics, because it is located on the fine line between human subject and human non-subject research. On the one hand, the data are deidentified, and were collected without intervention or interaction, while on the other hand, it concerns personal data on humans. To explore how register researchers approach the research ethics of working human data, insight into the normative framework that applies in these types of research is needed. A content analysis was conducted on academic codes of research conduct (see Chapter 3 for methodological details, Appendix 1 for the codebook, and Appendix 2 for an overview of contents), which yielded three human data research ethics values (*respect for persons, beneficence, and justice*), and one prerequisite factor (*competence*). Several applications of these values were identified, namely *consent, privacy, assessment of risks and benefits,* and *selection of subjects*. In what follows, the ethics values and practices will be discussed and defined. After, *competence* will be covered. Finally, the conceptual model employed in this study will be presented.

2.1 Respect for persons

Respect for persons requires a researcher to acknowledge a person's autonomy (Vereniging van Universiteiten, 2014), and to protect those with diminished autonomy (United States, 1978). In practice, this means that persons should be given the opportunity to enjoy their rights of self-determination, i.e. that they can decide for themselves what will happen to them (United States, 1978). To meet this requirement, research can only be conducted on people who wish to participate. Researchers should ask for the subjects' *consent*, which is defined as 'to explicitly agree to participate in the research' (United States, 1949). For the consent to be valid, it should be voluntary given, informed, and of sufficient quality (European Science Foundation and ALLEA, 2011; United States, 1949, 1978; Vereniging van Universiteiten, 2014). Voluntary consent means it is freely given, without coercion or undue influence, and it can be revoked at any time. Informed consent means that subjects make their decision about participation after they are informed about the details of the study, which should be presented in an understandable way. Consent is of sufficient quality when subjects have the legal capacity to consent, understand the risks and benefits related to the research, and understand the voluntary nature of participation. People who cannot give consent of sufficient quality should be protected by the researcher (United States, 1978).

Respecting a person's autonomy comes together with respecting their privacy (Vereniging van Universiteiten, 2014), which is defined as the different extents to which a subject can "determine whether and how information about [his/herself] is revealed to others" (Stein & Sinha, 2002), or remains shielded from public attention. Three research-related activities affect privacy, namely information collection, processing and dissemination (Solove, 2006). Some subjects participate on anonymous basis, where their name remains unpublished, or excluded from the data, while others want to be fully de-identified, where not only their name, but other identifiable information is removed as well. Those who consent to participation in research, share their information with a researcher, but might not want their personal information to be recognizable from a publication or the raw data. They might also have preferences with regard to the purposes for which their information can be used. Some might not want their information to be used in any other projects of the researcher, distributed among other researchers, and/or even to third parties. Confidentiality is the extent to which the information that the participant shares with the researcher remains between them, or is shared with others with permission of the subject (European Science Foundation and ALLEA, 2011; Solove, 2006). Confidential data treatment thus requires researchers (as well as data producers and providers) to follow the agreed-on characteristics of the research about who has access to which information of the subject, and under what conditions.

2.2 Beneficence

Beneficence is the obligation of researchers to take the welfare of participants into account (United States, 1978). In practice, this means they should avoid harming those involved in research, while they also should aim to maximize the possible benefits, and minimize possible risks that come with research participation (European Science Foundation and ALLEA, 2011; United States, 1949, 1978; Vereniging van Universiteiten, 2014). To fulfil this obligation, the researcher should assess potential benefits and risks of harm, which requires them to carefully think about the positive and negative consequences the research might have for the different parties involved (Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2010; United States, 1978). Benefits are positive influences on health, safety, or welfare, while risk is the possibility that harm occurs, indicating both the probability and severity of potential harm (United States, 1978).

2.3 Justice

Justice involves a fair distribution of benefits and burdens (United States, 1978) resulting from research, which can be achieved through a fair selection of research subjects. This requires researchers make conscious choices about who to include in, and exclude from participating in research, and to consider what the consequences of these choices are, in terms of distribution of positive and negative outcomes (European Science Foundation and ALLEA, 2011; United States, 1978). The outcomes of a selection procedure are fair when every individual has the same chance of participating in beneficial or risky research (individual justice) and when already burned groups of people (stigmatized or otherwise vulnerable groups) are not further burdened by research participation (societal justice) (United States, 1978).

2.4 Competence

We define *competence* as *possessing the required abilities, skills, and knowledge to conduct research effectively, and in an ethically correct manner*. Following from this definition, competence is a prerequisite to be able to adhere to the aforementioned ethical values. Researchers might have the best intentions with a research subject, but when they are not competent, reaching *beneficence*, *respect* and *justice* becomes problematic as they simply lack skills, knowledge, and/or abilities to succeed. To have competence, researchers need to be sufficiently qualified (United States, 1949), which means they have to actively maintain or improve their level of expertise by undertaking “relevant education, training, supervised experience, consultation, or study” (American Psychological Association, 2010:2.01c, comma added). This implies that researchers should conduct only research with populations, and in areas, where their competence is sufficient, that is; they are trained, have professional experience, or obtain supervised experience when they conduct their research projects (American Psychological Association, 2017). Moreover, to meet the requirement of competence, researchers should be aware of (inter)national laws, regulations and protocols related to research, and adhere them (European Science Foundation and ALLEA, 2011; Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2010). In practice, this may imply for example that the researcher includes systematic and comprehensive information about the risks and benefits of the research in a research proposal, which is then assessed by a review committee (i.e. institutional research board, ethics committee, research ethics board), to see whether it meets scientific and ethical standards (Dolgin, 2014).

2.5 Conceptual model

In the previous sections, the theoretical framework of this study was outlined, aiming to give insight in how register researchers approach the ethics of working with human data. Three human data research ethics values were identified, namely *respect for persons, beneficence,* and *justice*. *Competence* was defined as a prerequisite for bringing these values into practice: sufficient expertise and awareness of protocols helps researchers interpret the values, and form perceptions of what ethically sound register research looks like. The conceptual model displays how human data research ethics are interpreted by the researcher, who then forms a perception of which practices are important to ethically conduct research (see Figure 1). Chapter 5 presents an inductive version of the conceptual model, that was established based on the findings of this study.

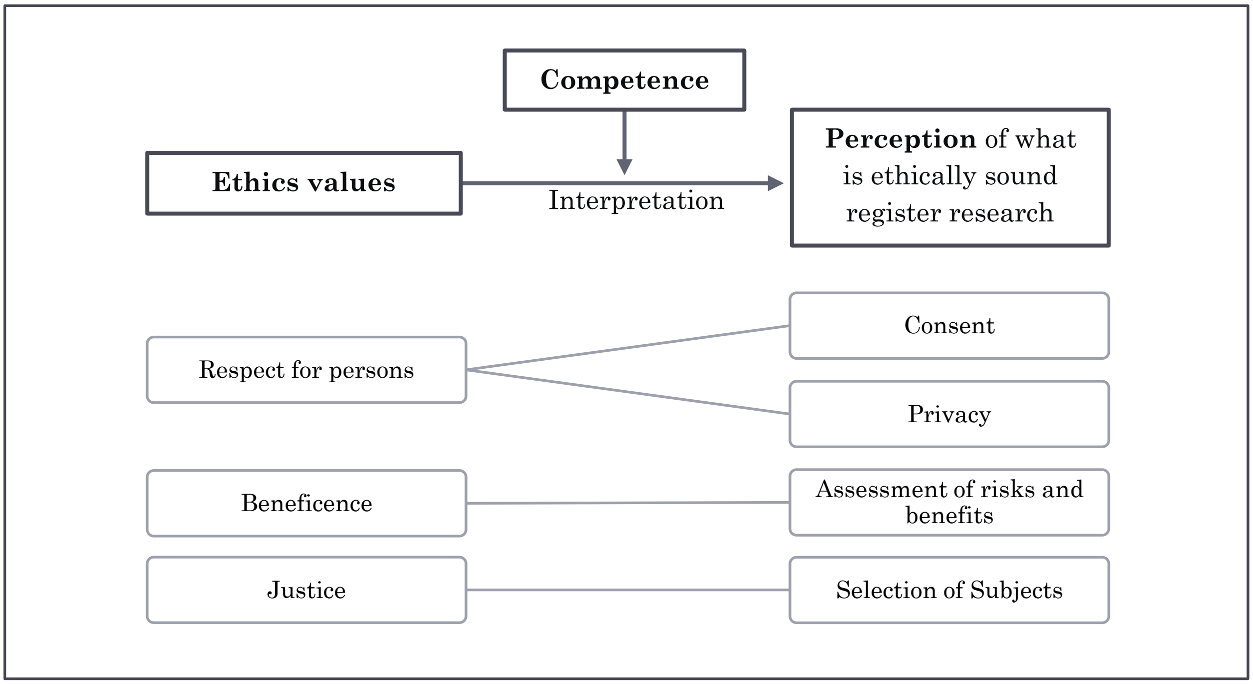


Figure 1 Conceptual model

CHAPTER 3: METHODOLOGY

In what follows, the design of the study, including data collection methods and an overview of the analysis will be described.

3.1 Study design and data collection methods

This project explored register researchers' perceptions and attitudes towards research ethics by employing a qualitative mixed method study design, where both a content analysis of existing documentation on research ethics, and a thematic analysis of semi-structured in-depth interview data, were combined to reach the objective of the study. A schematic overview of the study design is included in Figure 2.

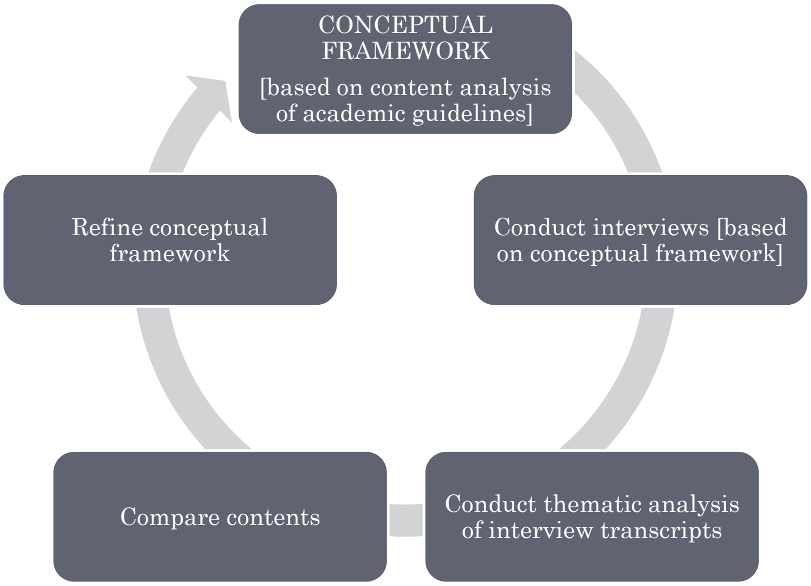


Figure 2 Study design

Some researchers consider qualitative research methodology crucial for an exploratory and in-depth understanding of ethical decision making, as it is an "interactive and iterative process of participants and researchers jointly exploring the phenomenon to yield rich insights for theory building" (Lehnert, Craft, Singh, & Park, 2016:499). Moreover, qualitative methodology helps researchers "*interpret* phenomena in terms of the *meanings* people bring to them" (Hennink et al., 2011:9). This study employs a qualitative *mixed* method research design, as mixed method research is thought to be an effective way to examine different aspects of the research problem (Hennink et al., 2011). We aim for a grounded theory approach, where we examine the linkages between data and theory, in order to build an inductive conceptual model which captures the theory developed from research findings (Babbie, 2010; Bazeley, 2013; Hennink et al., 2011; Maxwell, 2005).

As described in Chapter 1, there is little empirical background on the topic of ethics in human non-subject research. Conventional content analysis is a most suited research design for situations where “existing theory of research literature on a phenomenon is limited” (Hsieh & Shannon, 2005:1279). It helps to systematically make interferences with regard to manifest and latent meanings of messages, by sorting and linking codes derived from the data (Bazeley, 2013; Hsieh & Shannon, 2005; Saldaña, 2011). Although the conventional approach to content analysis is considered to be unsuitable to demonstrate relationships between concepts, it is a helpful tool in the development of conceptual models (Hsieh & Shannon, 2005).

In this study, content analysis was employed to gain insight into the guidelines that were developed for academic population research in the Dutch context. A large body of ethics guidelines and written protocols for research exists, aiming to set a normative framework in which researchers working with human data make decisions about ethics. To get a clear picture of the norms and customs that apply in research on human data, the contents of eight international, national, and institutional academic codes of research conduct were analysed. Content analysis of codes of research conduct has been done before by for example Komić et al. (2015) and Resnik, Rasmussen, and Kissling (2015). Such policies set and enforce ethical research standards, and examining their contents is a useful way to study the norms that apply in research (Resnik, Rasmussen, & Kissling, 2015). Identifying the central themes present in such documents gives insight into ethics concepts and practices which might be to more or lesser extent relevant for register research. Therefore, these documents can be used as a reference point to explore how researchers approach ethics in their daily work. The themes that emerged from the data were brought together in the conceptual framework presented in Chapter 2.

This study also explores researchers’ perceptions and attitudes about the concepts included in the framework. Moreover, it examines whether there are additional practices of perceived importance for register researchers, and what the role of researchers’ competence is in the process of forming these perceptions. Organizations and business literature has a history of measuring perceptions and attitudes to for example workspace ethics with *scales*, on which the respondent can indicate the extent to which they agree with a statement of situation (e.g. Dawson, 1997; Hunt, 1997; Singhapakdi, Kraft, Vitell, & Rallapalli, 1995). However, this approach does not enable participants to truly explain how they feel, as their answers can strictly be given within the framework the researcher sets for them (Perryer & Scott-Ladd, 2014).

Perceptions are inherently personal and subjective processes, in which humans interpret and organize sensations, in order to give meaning to them. This means an individual observes a situation or stimuli, and attaches a meaning to it, based on prior experiences (Pickens, 2005). Attitudes are defined as a persons’ overall evaluation of the objects in their environment (Eaton & Visser, 2008). Attitudes differ in strength, which means people find some attitudes more important than others. Multiple features together form the strength of an attitude, for example the certainty with which it is held, the amount of knowledge on which it is based, and the degree of to which a person finds the attitude personally important (Eaton & Visser, 2008).

Perceptions and attitudes are thus formed on sensations and objects in the persons' environment (Eaton & Visser, 2008; Pickens, 2005). Consequently, employing a multidimensional approach to ethics, where such contextual factors are taken into account (Perryer & Scott-Ladd, 2014), was necessary to gain a detailed understanding of how register researchers *perceive* these factors. Moreover, an *open* approach to research ethics is essential because it leaves researchers free to share their perceptions and attitudes without feeling too limited by the data collection methodology.

Following this line of thought, a semi-structured, in-depth, face-to-face interview approach was employed, where new topics can emerge, and probes are used to gain a detailed understanding of the participants' perceptions and attitudes, in a relatively safe setting (Hennink et al., 2011). Subsequently, a thematic analysis was applied, where themes in the data were identified by integrating codes that entail both content and meaning (Bazeley, 2013). This was done by listening to the audio recordings, reading the transcripts, and coding them, as well as through memo writing, which captured, facilitated, and stimulated analytic thinking (Maxwell, 2005). The findings of the analysis can be found in Chapter 4.

Finally, we combined the two approaches, and used the findings to develop an inductive theory (Babbie, 2010; Hennink et al., 2011). The conceptual model derived from the content analysis of guidelines was refined with the findings of the thematic analysis of interview data to form an *inductive conceptual model* which can be used to understand how register researchers working in the Netherlands approach and practice research ethics. This model is presented in Chapter 5.

3.2 Content analysis of academic codes of research conduct

To get insight into the normative framework that applies in research on human data, the contents of eight academic codes of ethical research conduct were analysed. In what follows, the process of data selection and analysis of these codes is described.

3.2.1 Description of selection of academic codes of research conduct

Following the nature of register research on human data, the analysed codes were either interdisciplinary, or focussed on population studies. As the scientific community tends to be largely internationally organized, not only institutional, and national, but also international codes were analysed to get insight into the norms that apply in research on human data. Both current and historical documents were used, and all codes were written in English. The international codes of conduct analysed in this study are the Nuremberg Code (United States, 1949), the Belmont Report (United States, 1978), the report Best Practices for Ensuring

Scientific Integrity and Preventing Misconduct (OECD Global Science Forum, 2007), the Singapore statement (Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2010), the Montreal Statement (Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2013) and the European Code of Conduct for Research Integrity (European Science Foundation and ALLEA, 2011). One national code was analysed in this study, namely the Netherlands Code of Conduct for Scientific Practice (Vereniging van Universiteiten, 2014), and one institutional guideline, the Regulations for the Protection of Academic Integrity (University of Groningen, 2012). The documents were retrieved online, after which they were submitted into qualitative data analysis and research software ATLAS.ti (version 7.5.7).

3.2.2 Content analysis of academic codes of research conduct

Employing ATLAS.ti software, a qualitative content analyses was conducted to discover the messages about human data research ethics in the documents (Friese, 2012). An inductive approach was necessary, as the focus of this study lied on human data research ethics, for which coding schemes had not yet been developed. Codes were applied to passages about humans, and human data-, human subject-, human non-subject- research ethics and practices, while more general parts (for example about plagiarism, fraud) were excluded from analysis. Admittedly, this selection is arbitrary, which is why more ambiguous passages were included in the study as well. Following coding, the process of categorizing revealed certain themes in the data, for example, how *voluntary consent, informed consent,* and *consent of sufficient quality* all fall under the value *respect for persons*. Arranging the codes thus helped to identify ethics values, and applications of research ethics that were deemed important in research on human data. Competence was initially interpreted as an ethics value, but after some reflection, it was included as a contextual concept. As it is difficult for researchers to follow the norms stated in the documents if they are not competent, competence was modelled as a prerequisite to bring the values into practice. Parallel to the analysis, a codebook was developed. The code tree displays (sub)themes, and examples of how codes were applied, and can be found in Appendix 1. Together with some additional literature, the values and practices were used to build the conceptual framework of this study, which can be found in detail in Chapter 2.

3.3 Thematic analysis of semi-structured in-depth interview data

To get insight into register researchers' perceptions and attitudes towards human data research ethics, the research ethics they themselves apply in their work, and the role of competence in their work, semi-structured in-depth interviews were conducted with register researchers working in the Netherlands. The following section gives insight into the research context, study population, process of data collection, and analysis of the interview data.

3.3.1 Research context

This study focusses on register researchers working in the Dutch context, because the Dutch register information tends to be of excellent quality. Moreover, around 95% of the national statistics are produced by Statistics Netherlands, which Eurostat's peer reviewers recognized to be one of the leading National Statistics Institutes in the world (O'Hanlon, Szép, & Smrekar, 2015). Especially notable are the innovative methods Statistics Netherlands implemented in their statistical programme, including 'big data' research and strong collaborations with universities (O'Hanlon et al., 2015). Therefore, the objective of the study focussed on the Dutch context, and the location of data collection was geographically limited to the Netherlands. The research and data collection was conducted by a Masters student from the University of Groningen, supervised by an assistant professor from the same university. Throughout the research project, the strategies for data collection and analysis, as well as written work, were refined with input from the supervisor.

3.3.2 Study population and participant recruitment

Following from the research objective, the target population of this study consists of researchers working with Dutch register data on humans, in order to understand behaviour and population dynamics. They work at public research institutions in the Netherlands, such as municipal or provincial research offices, academic or applied universities, or governmental research organizations such as Statistics Netherlands. This population was chosen because we wanted to focus on research professionals providing policy-support or contributions to scientific knowledge. Therefore, researchers working in amateur or commercial settings are excluded from the study population. Potential participants were recruited through websites of research institutions, initial academic contacts, and by snowball sampling. The participants were send an email (included in Appendix 3) with information about the study, and the question if they wanted to participate.

3.3.3 Method: In-depth interviews

The human data research ethics *values* (as were described in Chapter 2) were operationalized based on the findings from the content analysis of academic codes of research conduct. A detailed overview of the operationalisations can be found in Table 1. These operationalisations were integrated into an interview guide, which was used as a reference point during the interviews, to make sure all topics were discussed.

Table 1 Definitions and operationalisations of concepts

Concept	**Definition**	**Operationalization**
Respect for persons	To acknowledge a person's autonomy, and protect those with diminished autonomy (United States, 1978).	**Consent**; a person explicitly agrees to participate in research (United States, 1978). Should be valid, informed and of sufficient quality (European Science Foundation and ALLEA, 2011; United States, 1949, 1978; Vereniging van Universiteiten, 2014). **Privacy**; the different extents to which a person can "determine whether and how information about [them] is revealed to others" (Stein & Sinha, 2002).
Beneficence	Obligation to take welfare of those participating in research into account: do not harm, and maximize possible benefits and minimize possible risks that come with participation (United States, 1978).	**Assessment of risks and benefits** by the researcher, where s/he considers risks as the possibility harm occurs, and benefits as positive influences on the health, safety, or welfare for the participant and/or society (European Science Foundation and ALLEA, 2011; Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2010; United States, 1949, 1978; Vereniging van Universiteiten, 2014).
Justice	Achieving a fair distribution of benefits and burdens across subjects (United States, 1978).	**Selection of subjects** should be fair for the individual and for society; every individual has the same probability of participating in beneficial or risky research, and already burdened groups are not burdened further by research participation (European Science Foundation and ALLEA, 2011; United States, 1978).
Competence	Possessing the required abilities, skills, and knowledge (American Psychological Association, 2017) to conduct research effectively, and in an ethically correct manner.	Sufficient **expertise** (United States, 1949) and awareness and adherence to **research protocols** (European Science Foundation and ALLEA, 2011; Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2010).

Four parts can be distinguished in the guide, which can be found in full in Appendix 4. The first part consisted of several opening questions, aiming to build rapport (Babbie, 2010; Hennink et al., 2011), for example about what topics the researchers study, and how they obtain the data. In the second part, they were asked about the role of topics as privacy, consent, assessment of risks and benefits, and data selection in their work. These questions were asked to explore researchers' perceptions of human subject research ethics concepts, and their attitudes towards the role of these ethics in register research. The questions from this part were also used as a bridge to have researchers reflect on the ethics and practices they currently apply in their work. The third part consisted of questions about competence, for example about the role of additional training, and codes of research conduct in their work. These questions were asked to find out what the role of researchers' competence is in forming perceptions about ethics practices. The fourth and final part consisted of questions meant to break down rapport, for example, how their own ideas about research ethics had changed throughout their career.

The interview guide was piloted, which revealed some issues with the wording of questions (which were afterwards changed to more 'every day speech'-wording), and led to the removal of one question, which showed to be too broad and generally unclear for participants, while its topic turned out to be covered by the other questions as well.

3.3.4 Description of the interview process

The participants (four men, seven women), worked at universities, and municipal or (semi-) governmental research organizations in the Netherlands. They fulfilled, apart from research, a variety of tasks, such as collecting, processing, maintaining, and distributing register information. They were a broad mix of junior and senior researchers conducting research on topics such as family, (un)employment, social security, health(care), poverty, and debt. Due to participants' confidentiality concerns, a detailed table of participant characteristics has *not* been included in this report, nor will their information be displayed in relation to the quotations.

The data collection continued until the saturation level was reached, which happened after 11 interviews. The interviews were conducted in May/June 2017, and were, with one exception, conducted at the working location of the participants. The duration of the interviews was about 45 minutes on average, but because the consent process took some time as well, the total time spent with participants was closer to one hour. All interviews were conducted by the same interviewer, and conducted in Dutch. The interviews were audio recorded and verbatim transcribed in Dutch.

During the interviews, I, the interviewer, choose not to reveal my own perceptions and attitudes about register research and the ways research ethics are approached in register research, with exception of the pilot interview. In that interview, I revealed my enthusiasm about the data, and the extent to which it was available, to the participant. This first interview was unique in the ways the rapport that was established and the topics that were discussed as a result. It is one of the richest interviews in the set, when it comes to getting insight into researchers' perceptions and attitudes. However, for reproducibility reasons, the interview guide was followed more closely in later interviews, and own ideas were not- or to very limited extent- revealed during the data collection. Some participants specifically wanted to know my stance, or discuss the topic in a two-way conversation, which we then did after the interview ended.

3.3.5 Data analysis

The audio recordings of the interviews were verbatim transcribed to capture the essence of what participants said, in their own words and expressions. This was deemed important because it allowed to obtain an emic perceptive about the matter, which helped gaining a detailed understanding of how register researchers approach research ethics in their work.

The transcripts were coded using the software package ATLAS.ti 7.5.7 (Friese, 2012), and a combination of inductive and deducted coding was applied to capture how researchers reflect on the ethics they themselves apply in their work, how they reflect on the role of human subject research ethics in register research, and what role competence plays in forming perceptions about ethics practices. Some of the codes followed from the topics discussed from the interview guide, for example about consent and privacy. Others were *deductive* codes, brought up by the participants themselves. Table 2 shows an example of such a code. A number of codes only came to the surface after extensive reading and analysing (e.g. responsibility in assessment of risks and benefits, responsibility in distribution of risks and benefits). A complete overview of the codes applied to the transcripts can be found in Appendix 5.

Table 2 Example of a deductive code

Code family	Code	Quotation
Perceptions/attitudes about research ethics applied in own work	Purpose limitation ('Doelbinding')	*P: "I am also bound to purpose limitation, so I can't just do anything with the data. It has to fit the statistical policy of Statistics Netherlands, and I am not allowed to bring anything outside, outside of this secured data environment."*

In the process of forming code families, many themes and combinations of topics were explored. Since the aim of this analysis was to answer the different research questions, the process of categorizing the codes was mainly deductive. However, in the end, subgroups of codes were formed inductively within the larger code families (e.g. how the subtheme *understandings of privacy* was formed within the code family *perceptions and attitudes towards human data research ethics*). Arranging the codes into subgroups helped identifying the themes related to research ethics that were evident in the data. These themes will be discussed in more detail in Chapter 4.

3.4 Data quality and study limitations

This study employed a qualitative mixed method approach, where the content of existing documentation on research ethics, and the themes in semi-structured in-depth interview data were analysed to explore how register researchers approach the research ethics of working with human data. Both methods yielded valuable insights into the role of research ethics in register research, but there are also some limitations to the methods.

With regard to the content analysis, a clear limitation was that only academic codes were included in the analysis on which the conceptual framework was based. This selection is arbitrary, and had the criterium been broader, for example by including laws, and (semi-) governmental guidelines, some concepts would have been included in the initial framework from the start, which would have led to less alterations in the refining process. On the other hand, including codes of conduct from different organizational levels (academic institutions, and their national and international partnership organizations) has shown to yield a diverse overview on the matter. The conceptual framework turned out to be a fruitful way to examine research ethics in register research, despite only academic codes being used in the process of establishing the framework.

A limitation of the interview process, was that none of the interviewees worked for a provincial research organization. Consequently, the perspective of researchers working 'in between' local and national government was not included. However, the group of interviewees was diverse, and to large extent able to cover a full range of perceptions on research ethics.

The data collection and (content and thematic) analyses for this study were conducted by one person, which can be both a strength and a limitation of the study. On the one hand, less variation in for example interviewing style and coding strategies can lead to more consistent data collection and analysis, while on the other, it can be difficult to give insight in the choices that were made in the process of data collection and analysis. As a solution for this difficulty, the data collection and analysis processes were described in detail in the methodology section (3.2 & 3.2), and details on coding were included in the Appendices 1 and 5 for full transparency.

3.5 Ethical issues

The following section describes the confidentiality and consent process, and reflects on the researchers' positionality in relation to the participants and the research topic.

3.5.1 Confidentiality & consent

Before the start of the interview, the participants were talked though the research process. The ways the data would be collected (interview, audio-recorded), stored (in password protected file on personal computer, up to two years) and disseminated (in such a way the information cannot be traced back to the participant) were discussed. After, the participants were free to ask any questions about the project or the data collection. If they had no questions, or the questions had been answered, the participants were asked to sign the initial email they were invited with (which also included information about the data collection and confidentiality), to demonstrate their consent to participation. All participants signed the form, and none objected to being audio recorded.

Ethics and practices can be considered sensitive topics, and thus confidentiality was -as expected- quite the theme for some participants. One participant even asked me if I worked for the data provider, Statistics Netherlands, afraid I was going to check their work for misconduct. Many of the participants were fairly concerned about being identified from this report, either because of things they said, or in the case their workplace or their research niche would be mentioned. As a result, the workplaces of the participants, as well as the dates on which the interviews were conducted, will not be included in the report. One participant requested to approve of any quotes of them that would be used in this report, which was complied with.

3.5.2 Positionality

Before discussing the findings, I want to address my own positionality as interviewer and researcher in relation to the interviewees and their research.

Personally, I am in a split when it comes to register research. On the one hand, I, a Dutch citizen whose data are also included in many register datasets, hold autonomy to high value. I am interested in privacy issues, follow current events with regard to 'big data', and try to do everything in my power to conceal the data traces I leave behind. The Dutch government collecting and storing large amounts of information on its citizens, and allowing researchers to work with these data, is something that I am personally strictly opposed to. On the other hand, there is me as primarily quantitatively schooled researcher, certainly seeing the pros of register research. It seems like a dream to work with datasets that contain information on so many people, and to be able to apply innovative methods and study research topics that previously could not be studied because the data quality was lacking. Thus, register research also appeals to me, and I can imagine that the ways you approach research ethics, your own perceptions and attitudes, might change- become of secondary importance, in relation to the research you conduct.

This current research is in many ways different from the register research the participants conduct on a daily basis. Some of the participants held explicit and implicit opinions about my - qualitative – research, and the ways I dealt with ethics during the data collection. For example: when discussing the process of data collection and confidentiality, some participants seemed to want to rush this part, interrupting me and finishing my sentences. Other participants commented on how *qualitative* researchers working in certain fields overdo ethics, thereby drifting away from the purpose of the research. It is possible that my position as aspirant researcher applying qualitative methodology, in relation to the participants conducting quantitative research, may have had consequences for the ways researchers answered the questions. For example, after going through the whole informed consent process themselves, they might have felt they could not speak freely about consent issues in register research, or might have felt I could not relate because conduct qualitative research.

CHAPTER 4 FINDINGS

The findings of the content analysis of ethics guidelines, as well as the thematic analysis of semi-structured in-depth interview data, will be presented together as follows: In the first section, the meanings register researchers attach to human data research ethics practices (as described by codes of research conduct), and the role these practices play in their work, will be explored (4.1). After, researchers' reflections on the research ethics they apply in their work will be discussed (4.2). Then, the role of competence in forming perceptions about ethics practices will be covered (4.3).

4.1 Understandings of human data research ethics in register research

The findings reveal how register researchers' interpretations of human data research ethics are in certain cases distinct from how these ethics were described in the codes of research conduct. While some concepts seem to have a very limited role in register research, others remain equally important, albeit being interpreted differently than in the guidelines. In what follows, the understandings of the ethics values *respect for persons, beneficence, and justice* will be discussed in more detail.

4.1.1 The role of *respect for persons* in register research

In the Belmont Report, *respect for persons* is described as a value which requires researchers to treat individuals as "autonomous agents" (United States, 1978:4), and protect those with diminished autonomy. In practice, this means that persons involved in research should be given the opportunity to choose what shall (or shall not) happen to them. The role *respect for persons* plays in register research was explored by examining the perceptions and attitudes register researchers hold towards informed consent and privacy. Three themes emerged from the data, namely *responsibilities in the informed consent process, perceptions on the role of consent in register research,* and *understandings of privacy*.

Responsibilities in the informed consent process

Respecting a person's autonomy requires "that subjects enter into the research voluntarily and with adequate information" (United States, 1978:6). Providing the people included in research with understandable information about the study, so that they can make an "enlightened decision"(United States, 1949:1), is seen as the duty of the researcher (European Science Foundation and ALLEA, 2011; United States, 1949, 1978; Vereniging van Universiteiten, 2014). People involved in research should for example be informed about the nature, duration, and purpose of the study, the methods used, the ways data will be collected, and "all inconveniences and hazards reasonably to be expected" (United States, 1949:1) in relation to the study.

However, the findings show that researchers are divided when it comes to actively providing information about register research to the people included in the data.

Some researchers perceive notions of information and consent to be of importance in register research. One researcher said for example: *"Yeah, actually, everyone that is included in a study, should be informed."*. Another mentioned: *"[I] think that it is pretty important that the public is aware of that, that the data exists, and that it is used, and that they could influence this, if they would want to."*. Others, on the other hand, see a limited role for providing the public with information about register research. Some researchers interpret 'providing information about register research' as: 'people can reasonably expect that register research takes place'. This shifts the responsibility to inform people about research, away from the researcher, to the public itself. Some deemed providing information unnecessary because people are confronted regularly with media items covering statistics (Statistics Netherlands publishes news items on their website which get picked up by established news outlets on a regular basis), which would allow the public to deduce that register data are used to produce these results.

> *P: "[I] think that people who are interested in that, can also definitely find somewhere that it is being used for research. Because you often hear about Statistics Netherlands on the evening news, they come with statistics […] those are based on something of course."*

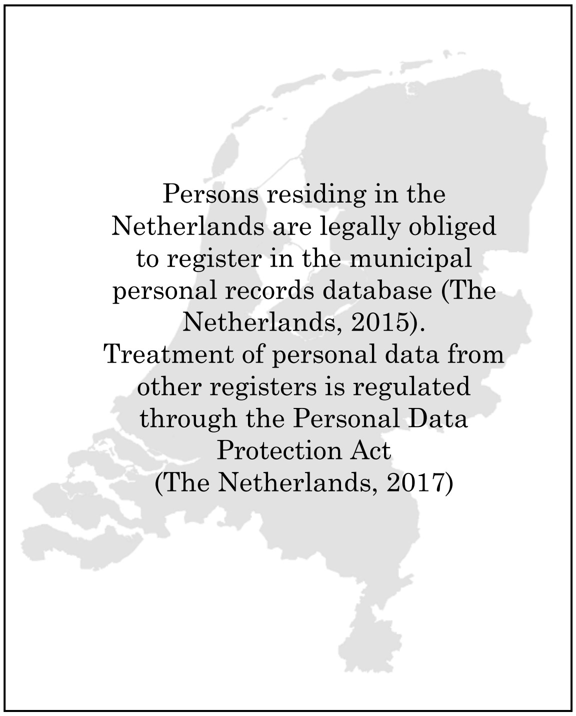
Whereas guidelines generally describe it as the duty of the researcher to provide people with information, some register researchers thus perceive it to be more the responsibility of the public itself, which demonstrates a shift of responsibility compared to what was set out in the conceptual framework. Some researchers oppose the idea of providing information all together. They fear commotion in society, if people were to find out register data are used in research. This commotion would mainly be due to a lack of insight and knowledge of the research process and safety standards. This researcher sketches how the public would react if they were asked to provide consent:

> *P: "The moment that you would really ask for consent, and I understand why that could be a theme- I think that you make people aware of something they don't want to be aware of. I think that if you were to tell people that the government has 20 years of residential information on them, they would be angry, but this is just necessary for the functioning of the country. I mean, you have to know how often people move, where they lived, for the statistics. And on the long run, you should always make the consideration: is the privacy of the individual more important, than the interests of society."*

Thus, while some researchers find it important that people are informed about their data being used in research, others see a limited role for themselves in this process. They experience shifted responsibility (from the researcher to the public itself) when it comes to providing information. Some researchers oppose actively providing information because they fear negative consequences, such as inadequate knowledge about the characteristics of the Dutch population.

Perceptions on the role of consent in register research

As described in the conceptual framework of this study, one way researchers can demonstrate their *respect for persons*, is by asking people for consent. Consent is an explicit statement in which the person agrees to participate in research, in voluntary manner, and based on sufficient information (European Science Foundation and ALLEA, 2011; United States, 1949, 1978; Vereniging van Universiteiten, 2014). Since register data were not primarily collected for research purposes, the role for consent is limited, as there is no direct contact between researcher and person whose data are included in the registers. Moreover, in the Netherlands, citizens have the *obligation* to register themselves (or else risk a fine) (The Netherlands, 2015), and the access to these data is regulated by law (see Figure 3). While Dutch citizens are expected to know the letter of the law, many might not know *which* data is included in the registers, or *how* these data are used for research.



Persons residing in the Netherlands are legally obliged to register in the municipal personal records database (The Netherlands, 2015).
Treatment of personal data from other registers is regulated through the Personal Data Protection Act (The Netherlands, 2017)

Figure 3 Illustration regulatory context

The Belmont Report presents examples of situations where, alike the Dutch context, individuals have limited self-determination. Cases of “illness, mental disability, or circumstances that severely restrict liberty” may require extra protection for individuals involved in research (United States, 1978:5). They present an example of a study involving prisoners who, on the one hand, should be able to exercise their right to self-determination, by for example making the choice to participate. On the other hand, they should be protected, as prison might have "coerced or unduly influenced [them] to engage in research activities", for which they would not have chosen had they not been incarcerated (United States, 1978:6).

Similar scenarios can take place when it comes to register research. Consider for example a situation where a person applies for welfare. On the one hand, they themselves hand over the necessary information for the approval process (which can be considered an active choice to share this information), on the other, this is not a choice to have their data included *research,* it is a choice made in order to obtain benefits (and avoid fines or lower allowances (The Netherlands, n.d.)). Thus, as register research is conducted in a context where people are not able to exercise autonomy (they are not *truly* free to decide what happens to them), register research *might* require extra protection for the people whose data are included. One researcher mentioned how citizens indeed have little control over what their information is used for in 'big data type' register research, and that "that is the reason the government makes those decisions for us".

Since the data are used for many purposes, and are included in many interlinked systems, researchers perceive when it comes to withdrawing from register research, due to the nature of the data and the way it is stored:

> *I: "Should it be possible for people to omit from research?"*
>
> *P: "Yes, absolutely! Yes, everybody has that right. […] if you work with a lot with big data, it just isn't possible, no, that is right. But I think it should be possible." [laughs]*

Another one said:

> *P: "[T]he system requires the integrity, so it forbids you from just removing things, […] It is a difficult predicament. On the one hand, I think: One should have the right to be forgotten, […]. But I'm afraid that is not technically feasible. But that would be the ideal situation for me, that you think: that is possible. If people would really want that, then they should have that right."*

Some researchers perceive notions of consent to be of little relevance in register research, because they interpret register research as mainly being macro research, that happens to be conducted on microdata. The research is not meant to follow individual cases, but to get insight into the larger picture. These procedures and characteristics of the research are by some researchers employed as motivations to not obtain consent. The idea is that, because researchers are not interested in the individuals 'behind the data', their role (and say) in the research process is limited. One researcher summarized the difficulties between the characteristics of register data research and consent issues as follows:

> *P: "And the difficult part is, in essence it's not about people the moment you do register data research, but all people are included in the data, and you could follow all people*

> *individually. And I think that is where the shoe pinches, you can talk about people, but […] that is not what I want. [S]o I think that [those] research guidelines aimed for individuals, informed consent, […] are not applicable per se. The ethical questions should focus on how to protect these data the best we can."*

To conclude, in register research, the role of consent is limited in practice. Registration is mandatory, and while some researchers feel that notions of consent are relevant and important in register research, they perceive constraints when it comes to removing people from the data. Others understand consent to be irrelevant in register research, as the focus of register research are not individual persons, but macro level phenomena.

Understandings of privacy

The analysed codes of research conduct provided no clear definition of privacy, and this inconclusiveness is to certain extent reflected in the perceptions of researchers as well. Table 3 illustrates the definition of privacy as was included in the conceptual framework, as well as the interpretations of privacy that are prevalent among register researchers. First, privacy is understood as *autonomy,* where people can exercise their rights to decide what happens with their information. Second, privacy is interpreted as *(non)identifiability,* a state where the information in the data is deidentified. Third, privacy is understood as *confidentiality*, where the information of the person who the information belongs to, remains between the person and the researcher.

Arguably, the role of privacy as autonomy is limited, because for many types of information, registration is a mandatory and mostly automated process, outlined in the law (See figure 5). Some researchers also perceive practical limitations to the autonomy interpretation of privacy, for example because they feel not sharing information becomes more and more difficult in the current information-focussed society, or because some situations require revealing information, that one might not *want* to share.

However, that data are collected and shared between (semi-)governmental organizations, and academic researchers, does not mean there is no role for privacy at all. Apart from autonomy, privacy is also interpreted as (non)identifiability, and confidentiality, where respectively the identity, and the information of persons remains shielded from public attention. Many researchers perceive privacy to be a combination of one or more of these understandings. Although they themselves have access to (to more or lesser extent) deidentified information, researchers utilize several strategies to make sure the information is not available for the whole world to see (see section 4.2).

Table 3 Understandings of privacy

In literature	***In conversations with participants***		
Privacy	*Privacy as autonomy*	*Privacy as (de)identifiability*	*Privacy as confidentiality*
People can decide *whether* they want to share information, with *whom*, and in what *ways*.	People can exercise their rights to decide what happens with their information.	The identity of the people in the data remains concealed.	The information of the people in the data remains concealed.
Extent to which a person can "determine whether and how information about [his/herself] is revealed to others" (Stein & Sinha, 2002)	*P: "I would say that […] privacy is, a combination of the right and the possibility to make sure that other people don't find out things about you that you don't want them to know about."*	*P: "[T]hat you can't retrace records on an individual level from the data. [T]hen I think the privacy is secured."*	*P: "But [the] privacy element is: as a researcher, you have information that you really have to secure and make sure isn't leaked."*

4.1.2 The role of Beneficence in register research

In the Belmont Report, *beneficence* was described as an ethical principle which requires researchers to protect subjects against harm, and also to consider potential loss of benefits initially gained from being involved with the research (United States, 1978). In this study, the role of *beneficence* was explored by looking at the perceptions register researchers hold towards assessment of risks and benefits. Three subthemes were identified, namely *perceived risks*, *perceived benefits*, and *responsibilities in the assessment process*.

Perceived risks and benefits

Conducting register research comes together with potential risks and benefits for individuals included in the datasets and for society as a whole. Most participants perceived the chance that risks occur in register research to be small. However, while the probability of risks of harm might (perceived to be) be small, does not imply that the risks are also thought of as small or neglectable by researchers. One participant described the risks and benefits as "two sides of the coin", where research can have both positive and negative externalities. Table 4 displays which potential risks and benefits researchers perceive to be relevant for register research.

Table 4 Perceived risks and benefits of register research

Perceived risks	*Perceived benefits*
• *Improper data use* • *Discrimination or stigmatization resulting from findings* • *Privacy violations*	• *Basis for custom policy* • *Meeting European Union requirements* • *Advance scientific knowledge* • *Cheap, quick, and less invasive way of data collection*

Responsibilities in the assessment process

The analysed codes of research conduct describe how these externalities are ‘weighed’ against each other, that is; risks should never outweigh the “humanitarian importance” of the problem under study (United States, 1949:1). Even then, the research might impose risks on those involved in research, while not yielding direct benefits for this group. In such cases, conducting the study might be justified, as long as the rights of those involved in research have been protected (United States, 1978). An important theme in assessment of risks and benefits is responsibility: who weighs these risks and benefits, and makes decisions about whether research can (and should) be conducted?

In codes of research conduct, it is perceived to be the duty of the researcher to assess risks and benefits inherent in their work (Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2010). However, many researchers, who use Statistics Netherlands’ data in their research, perceive the responsibility to make such an assessment, to have shifted from themselves to the data provider. One researcher explained for example: *“They [Statistics Netherlands] are always responsible for the ‘checks and balances’, so I think that […] that limits my responsibilities as a researcher as well, so to say.”*. Another shared a similar thought, where adhering the protocols shifts the requirement to assess risks and benefits from the researcher to Statistics Netherlands:

> *P: “Well, for us, the people that work with the Statistics Netherlands-data - it is assumed that Statistics Netherlands takes care of that. They collect the data, and they distribute it to us under certain terms, so that’s the principle. In that sense, they are the responsible party and we are just responsible for not breaking the rules.”*

Others do not experience this shift, for example because they work with municipal data, or are required to present the assessment in research proposals. Their perceptions on the necessity to weigh risks and benefits seemed to differ substantially, both within and across research organizations. Some researchers perceived the importance of assessment to be related to the aims of the research and the sensitivity of the data, while others considered it with the start of every new project.

> *P: "[Type of register research] only has few risks for individuals, in that case I don't really do it […]. But when the material is really sensitive, I think it should probably be done more often. […] Yes, it very much depends on whether it is sensitive or not."*
>
> *P: "[We] don't specifically take time […] to think about that, but there is always a consideration, every time a new assignment comes in we ask ourselves, yes, we do have that information, but what do they need that information for, and on what level do we provide, etcetera."*

Researchers approach the risks and benefits of register research as two sides of the coin, where the positive and negative outcomes of research are weighed against each other. Researchers working with data retrieved from Statistics Netherlands perceive the responsibility to assess risks and benefits of research to have shifted from themselves to the data provider. Those working with 'own' data hold various perceptions towards the necessity to assess risks and benefits.

4.1.3 The role of Justice in register research

The ethical principle of *justice* requires 'just' procedures for the selection of research subjects, so that positive and negative outcomes of research are fair distributed among those involved (United States, 1978). Selection can be just for individuals (fair selection in potential beneficial or risky research) and/or for society. Social justice of selection allows to make distinctions between classes of subjects, their ability to bear burdens, and "the appropriateness of placing further burdens on already burdened persons" (United States, 1978:18). Regarding social justice, it might thus be appropriate to prefer the participation of certain groups over that of others. In this study, the role of *justice* was explored by looking at the perceptions register researchers hold about the distribution of positive and negative outcomes resulting from the process of data selection and dissemination of findings.

Perceptions about distribution of outcomes

According to the European Code of Conduct for Research Integrity (European Science Foundation and ALLEA, 2011), researchers need to be sensitive to participants' characteristics, such as "age, gender, culture, religion, ethnic origin, and social class" (European Science Foundation and ALLEA, 2011:14, comma added). In the same code, scientists are seen as the responsible party when it comes to the choice of subjects, and the consequences of this choice. Researchers could for example exclude stigmatized groups to avoid burdening them even more, or publish results in such a way that the chance of stigmatization is minimal.

In register research however, this understanding of *justice* seems to be largely untenable. Researchers perceive data selection to mainly be a methodological issue, in which the aim of the research (which people are we targeting?), and the quality of the data (are their enough cases to draw solid conclusions?), together with some privacy considerations (can we publish it without violating privacy?) are the factors taken into account. When asked about whether distribution of positive and negative externalities of being included in research were considered in a particular research, one participant said resolute: *"No, because, because that's not the purpose of the research, absolutely, totally not."* Another researcher also mentioned how the aim of the research is leading in deciding whose data are included in research: *"That totally depends on the topic […]. In general, we just use everyone, so there is no selection."*.

Some also perceive the above-mentioned considerations about for example data quality to be leading in the dissemination process. One participant mentioned how they sometimes decide not to publish results because they were drawn on too small samples which would lead them to be statistically unsound, and prone to privacy violations.

> *P: "[I]f there is a risk for identification, that is an important criterion, or […] if one would suggest that with certain results a certain relation exists, […] while you have no idea, because the data quality was not sufficient- look, if the data quality is insufficient, then you don't state findings about that part."*

Conversely, other participants consider effects beside privacy violations in their dissemination process. Researchers mentioned for example how they conducted analyses with ethnicity as a variable (this is generally considered "*not done*", according to one of the participants), and not disseminating the results to avoid the backlash of ethnic profiling and discrimination. Some take into account how policymakers would interpret their work. They try to avoid misinterpretation of their results (which could lead to unfair treatment of individuals) by including disclaimers about the research in their publications. Others on the other hand, feel limited responsibility for possible unfair treatment of certain groups. One researcher explained for example how they think the results of register research can never directly lead to stigmatization:

> *P: "If you're talking about a group of people […], and you discover that the group in its entirety, […] does in general more of something that is not appreciated, then it's not a problem, I reckon. [L]ook, if people then stigmatize- then they all [fall for] the 'ecological fallacy' […]. The moment that people do that, yeah that's awful, but the research with register data is not the cause. [T]he cause is, human nature."*

Regardless of whether researchers pay attention to distribution of potential risks and benefits in their work, the general consensus seems to be that scientific research should be published. This is even the case if the results might not be favourable for certain groups, or might lead to severe responses in society. One researcher described it as follows:

> *P: "I reckon that actually: you should, and Statistics Netherlands does strive for that, […] you should communicate your research to the outside. […] [I cannot] withhold*

> *information because I think: oh, that doesn't really work for me, or that doesn't work for somebody else. [T]hat is the nice thing about science, you can just objectively show things, without directly having to deal with all kinds of subjective opinions. [laughs]"*

In conclusion, while some perceive methodological considerations to be leading in data selection and dissemination of results, privacy issues, and avoiding disproportionate burdening also play a role in the register researchers' work. With some exceptions -ethnicity-, most researchers feel that results of scientific research should be published.

4.2 Perceptions and attitudes towards research ethics

When it comes to research ethics, register research can be considered a stranger in our midst, balancing on the fine line between human subject and human non-subject research. In the following section, the perceptions and attitudes register researchers hold toward the research ethics they apply in their work will be discussed. After, researchers' understandings of enforcing ethics in a changing context are covered.

4.2.1 Approaching ethics

Register research involves analysing personal information of Dutch citizens on the microlevel, which comes together with ethical concerns about using data for other purposes than for which they were collected, data breaches, and reidentification. From the data emerged three strategies researchers utilize in their work to deal with these concerns, namely purpose limitation ('doelbinding'), separation of functions ('functiescheiding'), and de-identification. Table 5 shows a detailed overview of the approaches, and the meanings researchers give to them.

Many perceived the strategies to be an integral part of their work, and the analysis identified several motivations register researchers hold to approach research ethics this way. The researcher from the following quote explains what they do to avoid privacy violations, but also illustrates why they perceive this to be of importance:

> *P: "[W]e do everything we can, […] Not just because there is an ethical conscience, but also because Statistics Netherlands has [a central position]. That is so incredibly important, the trust that is put in us is so incredibly high, but it is also- yes, it took a long time to build that credibility […] You can […] very easily lose that trust. If you were to do something wrong. So it is incredibly important that we do everything we can to prevent the privacy from being violated […]."*

Table 5 Deductive approaches to research ethics in register research

Approach	**Meaning**	**Illustration**
Purpose limitation	Using the data that were provided for one specific purpose, that is: conduct a predefined analysis to answer the research question.	*P: "Statistics Netherlands provides the data, but they have a lot of terms for this[.] [Y]ou can only answer the questions that you have defined a priori in your research question. [A]nd you have to discuss those beforehand, […], then you can get the data, then you can receive the output you need to answer your research question, so you can't just search for interesting findings in the data, thus it is treated in a responsible manner."*
Separation of functions	Separating tasks within organizations or specifying the duration of tasks, to avoid that unauthorized persons access register data.	*P: "[E]specially here at Statistics Netherlands, it is very procedural, very strictly organized, with the different departments with different theme's[.] [Y]ou can't just access everything, you have restricted […] access to read and write data within the network, and those permissions are all temporary, so if a project has ended, then you also lost that. And it's all managed independent from the researcher. Permissions are granted and removed."*
De-identification	Removing identifying information from the data and output.	*P: "You have to make sure that records don't end up on the streets, you always publish in table format, so from a dataset, whether it's register data or surveys- you're never going to find an individual in there, and say something about them. And the table format is always made in such a way, that identifying people is not possible."*

In the above-mentioned quote, the researcher holds the *value* that identifiability should be avoided, but also mentions the position of Statistics Netherlands, the organisation they work for. Protecting the *image* of Statistics Netherlands as trustworthy party, who will do everything in their power to avoid that information can be compromised, also acts as a motivation to minimize the risk of privacy violations. Other researchers shared their thoughts about motivations to avoid these violations as well, and their reasons tend to be just as divers. For example, for some, avoiding this type of risk seems to be integral part of what some researchers consider 'being a good researcher', for example:

> *P: "[The] quality control, that you have to be able to reproduce your findings, that you don't plagiarize, that you handle personal information with care. It is […] like with a doctor, you expect a researcher to be good at their job, and those are all important facets of being a good researcher. Yes, the first time I did think it was a bit gibberish, but now I am convinced that it is actually important. Yeah, especially if you are not treating it with care, you get commotion in society, if there is abuse-. And at a certain moment […] measures will be taken, and I think that would really be a shame."*

In this case, the researcher has a specific idea of how a 'good' researcher behaves, and actively tries to comply with this idea. Wanting to adhere to this image is itself a motivation to avoid privacy violations. Moreover, the participant is afraid violations of privacy might lead to commotion in society, which could lead to researchers being denied access to register data. Similarly, the researcher from the following quote mentions the possible repercussions that could follow from privacy violations as a motivation to avoid them:

> *P: "It is important: […] how do you communicate things, how do you report things? It should be in such a way that you can't trace things […] on the individual level. Then, I reckon that you violate privacy, and then as an organization […] you can get sued. But we handle it well."*

Thus, while many researchers perceive purpose limitation, separation of functions and de-identification as important and reliable strategies to avoid privacy violations, their motivations to apply these strategies differ considerably. Some are concerned with their own or their organizations' image, whereas others mainly want to avoid public uproar or sanctions.

4.2.2 Approaching ethics in a changing context

The Dutch Data Protection Authority ('Autoriteit Persoonsgegevens') has been given the task to supervise and advise organizations and institutions working with personal data. In principle, every research institution where personal data is processed, should follow privacy regulations and can be audited by the Data Protection Authority. Previously, the Authority could only impose administrative sanctions, however, since 2016, it has the power to administer financial penalties as well (Autoriteit Persoonsgegevens, n.d., 2015). Along with the change in sanction structure, came the obligation for companies and (local) government to report (severe) data breaches to the Authority.

This change in regulations resulted in researchers approaching ethics in a different way. On the one hand, they feel their institutions took research ethics more serious, ever since fines could be administered, but they also mention how they are themselves more aware of rules and regulations. One researcher said:

> *P: "[T]hat data leak hotline, that alerted a lot of companies like oh, this could cost us a lot of money. You know what it's like: if it's going to cost money then we wake up. While these are all things of which you think: yeah, we should have always done this, and all of a sudden, they are serious. That is actually ridiculous, that it's always been like that. But suddenly we received training and awareness sessions and everything and then, for the most you think: "yeah, I already know this", but sometimes there's something that makes you think: "right, […] he actually has a point". And that makes you think a bit more critical about your actions."*

The previous quote illustrates how the researcher felt that enforcing ethics through fines resulted in more attention to research practices in their own, and other research organizations. Other researchers did not perceive the administrative and financial sanctions to be of much relevance, but feared bad publicity for themselves and their research organization if the Data Protection Authority would conduct an audit and present their findings. One researcher commented:

> *P: "[T]he worst that can happen is obviously that you will be publicly pilloried…"*
>
> *I: "To whom will this happen?"*
>
> *P: "[T]he researchers, the [organization]. But that is actually the only sanction that is enforced. […] [I]t is actually pretty interesting, because they visited [another organization] ... We were happy that they did not visit us, because, well, we would not want to be audited ourselves. [Other organization] pretty much do[es] the same things that we do, and, well, I am very curious how the privacy watchdog will react."*

The presence of the Date Protection Authority, and the possibility for repercussions makes researchers more aware of their own research practices. On the one hand, this leads to pro-active changes in how research ethics are approached, on the other hand, to 'wait and see' attitudes, where results of audits conducted elsewhere are used as a reflection on own research practices.

Enforcing ethics might be even more important in the changing context in which researchers work. The combining or 'matching' of datasets to population register- or other register data sources is an increasingly more common practice in the field (Willenborg & Heerschap, 2010). One researcher commented: *"You can use it to enrich your data",* while another commented how it allows researchers to study phenomena which could not be studied to such extent before. However, these developments also lead to concerns when it comes to data safety. One researcher said for example: *"Right, the risk is just bigger, because you are combining more. Because there is more data, there is more risk."*. Figure 4 demonstrates how researchers perceive this change to be approached in two conflicting ways, namely increased and decreased administrative control.

Increased administrative control	**Decreased administrative control**
P: "[I] reckon that everyone does realize that we obtain increasingly more data, and that there is some work to be done in terms of coordination and storage. Something we encounter often nowadays, is […] an audit, that you have published an article, and the you have to be able to reproduce your tables, with the help of existing data files, but then they also look at how you saved your file […]."	*P: [Y]ou notice that the access to data is given faster, so for example with [type of] records, well, nowadays that's a simple call, so to say, and then I just have to put it on paper, but in principle I could have those at my disposal. [A]nd that used to be much harder. So in the past […] the board had to make that decision. […] So the ease with which you can obtain data, that is increasing."*

Figure 4 Conflicting approaches to increased availability of data

As shown in Figure 4, researchers perceive the increased availability of data to lead to new infrastructure and stricter organizational policies regarding data management, while others feel obtaining access to data has become much easier, since it can now be requested without formal administrative processes. This development results thus in to two opposite reactions, where administrative control in organizations gets either increased (strict data management) or decreased (moderate control on data access).

Most researchers perceive the increased availability of data, and possibilities to combine data sources as a positive development, where they feel the positive outcomes of research abundantly exceed the disadvantages. Others, on the other hand, personally oppose the process, which can be difficult as they also benefit from it. Table 6 illustrates through a case study how one researcher perceived the changing context.

Table 6 Case study of a register researcher

One participant has worked for many years as a statistical researcher. During their career, the researcher has seen the field of register research change. Of course, register data were always collected and available for them to analyze, but the number of different datasets has increased over time, and has become more easily available. On the one hand, the researcher is content with this development: the data are certainly interesting, and allow to study many different research topics into great detail. On the other hand, it also worries them. Data gathered under the premises of being erased after a certain period, turned up, and was offered to the researcher. This lead them to conclude that collected information will never be discarded. The researcher actively removed themselves from voluntary registers, because "[…] I know what I can do with it! [laughs] I know for sure that it will never be deleted!". *The researcher opposes the data sharing, almost feels like a victim, but also plays an active part in the process. This puts them in a difficult position, which they feel is confrontational to talk about. The researcher wonders whether the sharing of data is a fait accompli. "Time will tell..."*

In conclusion, changing contexts also change the ways researchers approach research ethics. Increases in formal enforcement of research ethics provides researchers and their organizations with incentives to pay attention to research practices and to adhere protocols. Researchers feel register data to be increasingly available, which they think also leads to increased risks for data safety. Organizations deal with increased data availability through increased and decreased administrative control. For researchers who oppose the development of data sharing and matching, the changing context sometimes results in inner conflict, where researchers feel divided between their own values on the one hand, and the possible positive outcomes of register research on the other.

4.3 The role of researchers' competence in forming perceptions on research ethics practices

Several guidelines mention elements related to *competence,* which describes a situation where researchers have sufficient qualifications and awareness to protocols, to conduct research effectively and in an ethically correct manner (American Psychological Association, 2017; European Science Foundation and ALLEA, 2011; Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2013; United States, 1949). This requires the researcher to actively maintain or improve their level of expertise, while also being aware and adhering to relevant laws, regulations, and research protocols. This section covers the role of competence in register research, by examining what researchers perceive to be of importance when keeping their skillset up to standards, and by finding out what meanings researchers attach to the written protocols when making decisions in their work.

4.3.1 Improving and maintaining expertise

When it comes to maintaining or improving the level of expertise, three strategies are employed, namely institutional learning, self-study, and learning in the work environment. Some researchers keep their expertise up to standards by conducting literature studies, reading about methodology, and repeating material. One researcher said for example: *"[I am] still busy with that, so I do read things, repeat, like: right, how does this work again, how did that work again."*. Others also involve institutional learning in their learning strategies, for example by following additional education or courses though academic or other professional institutions. One participant mentioned:

> *P: "[I] followed [a couple] of courses at the university last year, so in that way I am involved in supplementary training. And for studies that are new, that are set up 'from scratch', we do literature studies, so by doing that I try to keep up."*

Not only academic institutions, but the data provider (Statistics Netherlands) plays an active role in improving knowledge and skillsets as well. They provide for example regular newsletters involving datasets, methodology, and research practices, and organize user meetings that provide a platform for researchers to talk about these topics and their own work. Researchers perceived the attention to research ethics to be limited on such days, although they also mentioned how some events were specifically organized to discuss research practises and ethics. However, some researchers perceive such events to be of little importance, because they feel research ethics is an established field, where new information comes up scarcely. One researcher commented: *"[Organization] organizes some lectures about that, but yeah, honestly: that doesn't really bother me that much because yeah, there is never really something new that they talk about. Nothing that hasn't been talked about before."*. Even though research ethics might not be considered the most innovative topic by some, Statistics Netherlands found new ways to draw attention to the topic: they ask researchers to answer a multiple-choice question when logging in to the secured data environment, to remind researchers of good research practices. Researchers perceive these questions to be a fun and helpful way to keep their knowledge up to standards, and to help educate new researchers on the topic. One participant said for example:

> *P: "Pretty funny, at Statistics Netherlands they actually have, before you log in, you have to answer a question, and that is always about a code of conduct […]. And if your answer is wrong you cannot log in. So, in that way they keep it topical, the knowledge. [I] think that is actually pretty creative, yeah and, I also think that it helps with the education of new researchers."*

The work environment itself is another way researchers maintain and improve their expertise. Participants mention how they keep their skillset up to date by regularly working with register data, participating in on-the-job training sessions and discussing research practices with colleagues. Discussing these practices with colleagues helps researchers in two ways. On the one hand, talking to non-expert or junior researchers presents the participants with situations they never experienced before, forcing them to think about research practices, while on the other hand also keeping their own knowledge up to date. One researcher mentioned:

> *P: "[E]very time a new researcher arrives, then you have to tell them: we treat data like this and that, you can do this, you can't do that, this is how you should approach it. This is the method that we use. That is the nice thing about PhD students, that, you do that together, and that you yourself also stay sharp, then it also can't deteriorate."*

On the other hand, discussing research practises with experienced colleagues makes researchers aware of a *habitus*. This habitus is a working culture where certain research practices are common, and where most colleagues work in that specific way. One researcher said: "*[…] by colleagues, that also work in a specific way, you copy that, so there is something akin to a culture in there, and a certain way of working that fits there."*. The habitus helps researchers decide which research practices to apply, while also giving direction on which practices to avoid. One participant said about this: *"[I]t is a kind of habitus so to say, that we just have here, that you just can't do certain things, but you also won't do them."*. Habitus is by some described as a self-regulating mechanism, where discussing and working together with colleagues keeps researchers on the right track. Researchers said for example:

> *P: "[The rules] they are something that you have to be aware of all the time, […] that is something that makes a good researcher, but it's also somewhat hard, because you have to safeguard that together."*
>
> *P: "[Navigating] is something that you have to do yourself, together with your project- the people within that team. With whom you collaborate in research. There is no one that actively monitors you, no, not like that."*

Researchers thus apply different strategies to improve and maintain their expertise, which helps them keeping their knowledge up to date and conform to the habitus at their working place. This habitus is to an extent a self-enforcing mechanism, where talking and discussing research practices with colleagues helps to keep researchers on the right path.

4.3.2 Giving meaning to written protocols

Many written protocols are established to guide (register) researchers in their work. Not only academic guidelines (e.g. European Science Foundation and ALLEA, 2011; Vereniging van Universiteiten, 2004, which were employed to build the conceptual framework), but also other institutional guidelines (for example the remote access protocol from Statistics Netherlands (Centraal Bureau voor de Statistiek, n.d.-b), or the guideline from collaborating research organizations in the Netherlands (MarktOnderzoekAssociatie, Vereniging voor Beleidsonderzoek, & Vereniging voor Statistiek en Onderzoek, 2010)), and laws (see also Figure 3), aim to help researchers make decisions in their work. The meaning researchers attach to these codes differ substantially, and will be discussed according to the metaphors included in Figure 5.

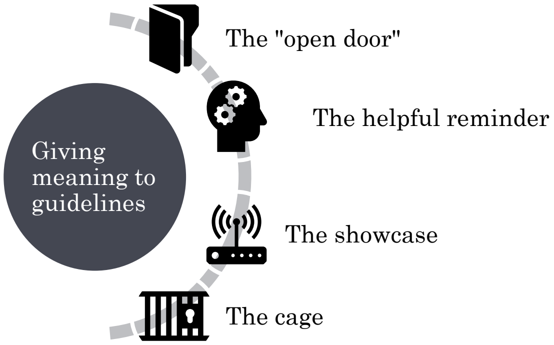


Figure 5 Giving meaning to guidelines through metaphors

All interviewed researchers were aware of the existence of protocols, though many admitted to not being able to reproduce the contents. One researcher said for example: *"Yeah, I actually should know all of this, but I don't."*, while another one mentioned: *"Yes… I honestly have to tell you that I haven't really looked into laws or anything. Yeah, I've looked into it but, there isn't really much in there."*. This last quote also illustrates how some researchers perceive the protocols to have predictable content. The contents are frequently described as an "open door" (a Dutch expression which means 'stating the obvious'), where the codes are interpreted as being "logical", "simple", "straight forward", or "common sense".

> *P: "Yeah, that sounds logical to me, […] It is good that they put it on paper, that you sign those and that you are aware of them, [they are things] that actually, if you just, think rationally, then you could think of them yourself."*

However, that the content is considered to be predictable, does not imply that the codes have little meaning to the researchers. Some perceive the documents to be helpful reminders in their work, because they emphasize the importance of research ethics. Many researchers mention how they have internalized the contents of the documents since they started working. As a result, the documents are not consulted on regular basis, but are sometimes used as work of reference during the research process and staff evaluations. One researcher commented on the *reminder* role of guidelines:

> *P: "Well, […] the funny part is actually: if you read the code, then I always think: yes, obviously. But it is just a reminder for everyone like: how should you handle your data, and what are you doing as a research, so what is the goal of what you are doing. And often it is pretty good to realize that: right, I am working with the data of other people and should treat that with care […]."*

Some researchers do not consult the protocols at all. They perceive the guidelines to be a showcase mechanism, where the contents are meant to show the public how organizations make sure the research is conducted in an ethical manner. In this interpretation, codes of conduct are thus not meant to be utilized by the researcher, but serve as a showcase for *credibility* for researchers and their research institutions. One participant said about this:

> *P: "[I]t is also important for Statistics Netherlands, to have something [guidelines] like that, to show the public that: we have this sort of rules, for the researchers that handle personal data. If anyone starts asking questions, then they can tell them that there are very clear rules. [So] on the one hand everything is made clear, on the other hand it's also the 'credibility'. So they can say: we do that-and-that-and-that."*

Some researchers perceive the protocols to be unfitting or too restraining to apply in their work, an interpretation that can be described as the *cage*. This is for example the case in situations where researchers can only work locally with the data, leading to situations where workstations are scarce, and researchers cannot work from home. Other instances are situations where researchers want to discuss findings with supervisors, but have to pay and wait for their research output the be approved by Statistics Netherlands. One researcher mentioned for example how they perceived the local workstations to be a constraining factor in conducting research, and explained how they found a way around it:

> *P: "[I]t is, practically speaking, very inconvenient when you have to run to a screen every time, so what we did was: we made a screenshot, and then we printed that, and looked at it together, and then make sure we'd destroy that print. So there are some elements in the procedure of Statistics Netherlands that make it very hard to do everything by the book and at the same time on a practical operable manner…"*

Another researcher shared a similar story, and commented:

> *P: "[A]lternatively […] for example when we sit at my supervisor, we have a meeting, and then I say: we are now at this-and-that table, will you walk with me? Then they can stand behind my computer, then they can watch, and subsequently… etcetera. That would be the alternative, and I think that alternative is not workable. [H]ow it is written on paper, [that is] obsolete in reality, the paper truth, and the paper norms and values, they just don't correspond with what is possible in practice."*

Clearly, protocols that were initially meant to guide towards ethical behaviour and 'safe' data treatment, can have unintended effects. These effects (screenshotting, printing through communal printers, unapproved output) are unwanted by both the data provider and the researchers themselves, but take place because researchers prefer some level of efficiency above following the protocols. Some researchers perceive certain guidelines to be inefficient, but decide to follow them anyway. One participant mentioned for example how some protocols are designed to appear good and useful, while in practice, they have barely any effect.

> *P: "[T]hat they force you to save social security numbers with encryption, so it's no longer a social security number. But it is still a uniquely identifiable number. […] Yeah, but those kind of measures, that is somewhat like window dressing. [I]t is about me, handling it properly. That is eh, and whether it's a social security number or an encrypted social security number, that is not relevant for me. That is completely irrelevant[.]"*

Other researchers experienced inconsistencies (research institutions or journals requesting copies of the datasets, requiring storage in certain formats) or grey areas were the protocols were not sufficiently guiding. Some researchers for example thought that the codes lack clearance when it comes to the increased availability of 'matchable' data. While some participants felt this should be solved by adding information concerning availability and matching of data to the guidelines (or other internal documents), others perceive briefer codes to be more effective, and argue against extending the guidelines.

Chapter 5 Discussion and conclusion

5.1 Synthesis of findings

The aims of this study were twofold. The first objective was to gain insight into the guidelines that were developed for academic population research in the Dutch context. Under this objective, we studied what the normative framework set by academic codes of research conduct consisted of. This was done by employing a qualitative approach, in which the contents of international, national and institutional academic codes of research conduct were analysed. We found three ethics values that were deemed important in research on human data, namely *respect for persons, beneficence,* and *justice*. Moreover, four corresponding research ethics practices were identified, namely consent and privacy (for *respect for persons*), assessment of risks and benefits (for *beneficence*), and selection of subjects (for *justice*). These findings were in line with existing literature, as the codes indeed aimed to guide researchers by distinguishing between desirable and undesirable practices (Komić et al., 2015; Koninklijke Nederlandse Akademie van Wetenschappen, 2012; Pimple, 2002; Resnik, Neal, et al., 2015; Resnik, Rasmussen, & Kissling, 2015; Vereniging van Universiteiten, 2014).

The second objective of the study was to explore how register researchers working in the Netherlands approach the research ethics of working with human data. A total of three research questions were raised under this objective.

First, we wanted to find out what meanings register researchers attach to human data research ethics practices as described in academic ethics guidelines, and what they perceive to be the role of these practices in register research. This was studied by employing a qualitative mixed method approach, in which the findings of the above-mentioned content analysis of codes of research conduct were combined with those of a thematic analysis of semi-structured in-depth interviews with register researchers. The findings reveal how register researcher's understandings of human data research ethics relate to their descriptions in academic codes of research conduct. Differences expressed themselves for example through shifts in responsibilities related to ethics, or practical limitations to apply the ethics in their daily work. As a result, researchers perceive some of the ethics practices that where described in the codes of research conduct to have a very limited role in register research. These findings relate to earlier ideas: academic codes of conduct do not explicitly focus on human non-subject research ethics, which *could* lead to situations where codes are unsuitable to apply, or where non-subjects rights are insufficiently protected (Rothstein, 2010). In practice, researchers avert this by strictly avoiding privacy violations, alike what was argued by Brothers & Wright Clayton (2010).

Second, we wanted to find out what register researchers' perceptions and attitudes are towards the ethics practices they apply in their work. This question was answered through the aforementioned thematic analysis of interview data. The findings reveal how researchers perceive avoiding privacy violations to be of great importance, and how they employ several strategies to avoid such violations, namely purpose limitation, separation of functions, and de-identification. This is in line with the human non-subject literature, where protecting privacy of the persons in the data is described as the only factor that separates human non-subject research from research on human subjects (Brothers & Wright Clayton, 2010). Researchers have various motivations to apply these practices, and experience changes in how they approach ethics as a result of perceived increases in availability of data, and formal enforcement of regulations. This experienced change of context is relates to the 'social media study' (Israel, 2015) that was described in Chapter 1.

Third, we considered the role of researchers' competence in forming perceptions and attitudes about research ethics practices. This was also explored by means of the preceding thematic analysis of transcripts from interviews with register researchers. The findings show how researchers employ different strategies to improve or maintain their expertise, namely self-study, institutional learning, and learning at the workplace. This is in line with the idea that researchers are indeed concerned with their qualifications (American Psychological Association, 2017). Especially learning at the workplace, where researchers learn with and from colleagues, is perceived a valuable way to improve expertise, as it creates self-regulating work culture which guides the researchers' behaviour. Researchers varied in perceived awareness of codes, and attached various meanings to the documents. Some perceive the protocols to be too constraining (alike described in (Israel, 2015)), leading to situations where researchers 'break out' of this 'cage', by not following the codes to the letter. This finding is opposite from what one would expect from literature (European Science Foundation and ALLEA, 2011; Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2010).

5.2 Towards an inductive understanding

The findings of this study show how register researchers approach the research ethics of working with human data. While the normative framework set out in the academic codes of conduct was generally found to be a suitable way to explore perceptions towards research ethics practices in register research, there were also some differences that should be considered. The findings of the study were processed into an inductive conceptual framework, which can be found in Figure 6.

Obtaining consent for example, is perceived by some researchers to be irrelevant in register research, while others do find it important, but experience constraints withholding them to apply notions of consent in their own work. When it comes to providing information about register research to the public, researchers are divided as well, but none take active measures to inform the public. Clearly, researchers perceive ambivalence related to both providing information, and obtaining consent, making it a concept that is considered of importance by some, but has very limited relevance in the daily work of the register researcher.

Researchers approach the value *respect for persons* mainly as protecting privacy. While privacy was initially understood as *the autonomy to decide whether to share information, with whom, and in what ways,* the relevance of this understanding of privacy turned out to be limited in register research. Register researchers attach two additional meanings to privacy, namely (de)identifiability, and confidentiality, where respectively the identity, and the information of the people in the data remain concealed. In practice, the relevance of (de)identifiability and confidentiality shows that protecting privacy is considered a VIP (*Very Important Practice*) by researchers, leading them to apply strategies such as purpose limitation, separation of tasks, and de-identification.

Researchers approach *beneficence* through assessment of risks and benefits. Risks and benefits are interpreted as two sides of the coin, and several examples of each were mentioned. The process of assessment was initially described to be the duty of the researcher, in which they weigh the potential benefits and risks of harm. However, in the daily practice of the researcher, this statement needs some refining. While some researchers indeed assess risks and benefits when they conduct their research projects, others perceive this to be mainly the duty of the data provider (Statistics Netherlands). Clearly, there is a role for *beneficence* in register research, but not all register researchers approach it by assessing risks and benefits themselves.

Whereas *justice* was initially described as fair selection of data, researchers perceive it to be less related to selection of data, and more to dissemination of results. Generally, researchers analyse the whole population (or select randomly) and perceive dissemination of scientific findings of importance. However, they refrain from publishing if they consider the findings to be methodologically unsound or prone to privacy violations. Some researchers take policy-implications into account when disseminating their work, to avoid unfair treatment as result of misinterpretation of findings. Thus, there is a role for *justice* in register research, but researchers approach it through dissemination of findings rather than selection of data.

Competence was indeed found to have a role in how researchers form their perceptions about research ethics practices, where researchers perceived maintaining or improving expertise to be a useful strategy to learn about research ethics. The role of protocols was understood in various ways, ranging from helpful to even counterproductive in achieving ethically sound research. These findings are especially helpful notions for organizations and policy makers, as they provide insight in what helps (e.g. learning with and from colleagues) and what does not (too many or too strict regulations, which will be bypassed under the guise of efficiency). Moreover, they present those concerned establishing academic guidelines with a clear overview of which elements, that are currently missing in guidelines, are relevant in human non-subject research, providing insight in how ethics are approached by those working with human non-subject data.

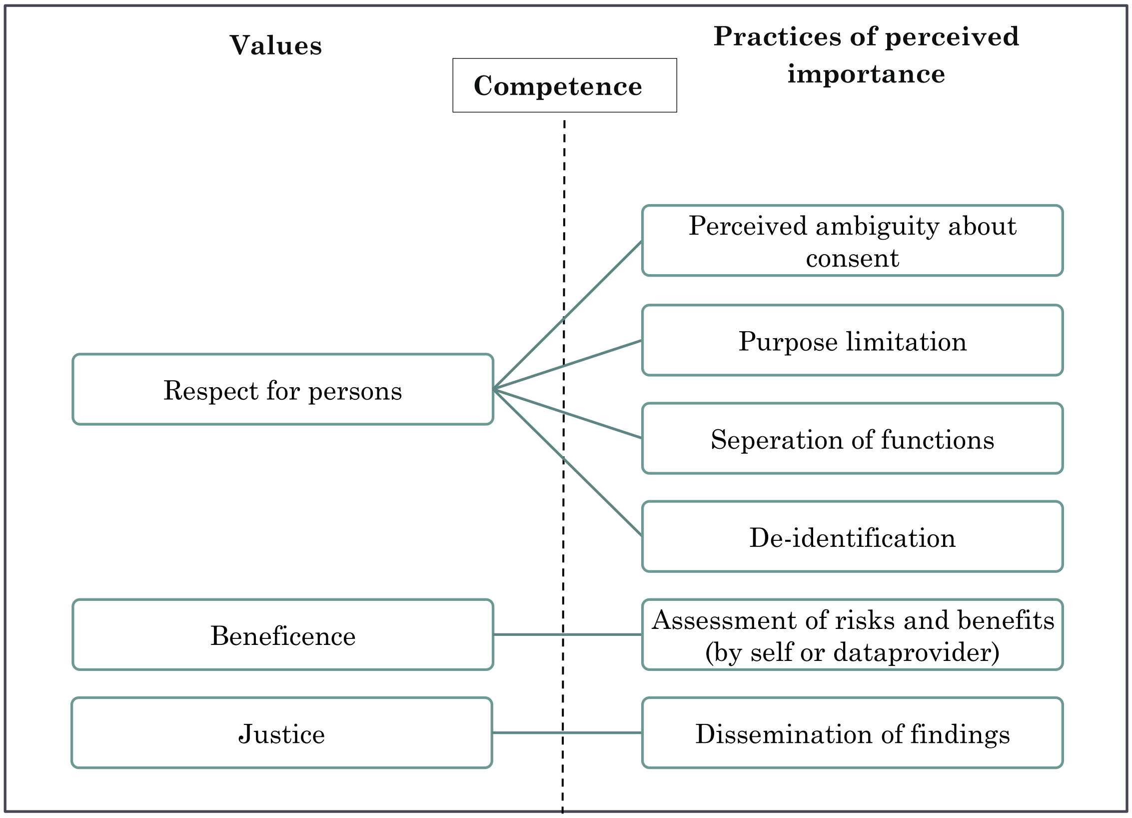


Figure 6 Inductive conceptual model

5.3 Limitations and recommendations

Having done this study in the Dutch context, we found both differences and similarities between how ethics are approached in the normative framework set for academic researchers by codes of research conduct, and how register researchers approach ethics in practice. A limitation of this study was the lack of empirical background on the topic, as literature on research ethics mainly focusses on human subject research, and not on human non-subject research. Moreover, there is, as far as we know, no existing literature on how researchers working with human non-subject data *approach* research ethics in their work, which made it difficult to embed the study in a body of scientific literature, or compare the findings to previous results. To work around this difficulty, we examined what *was* known about the research ethics of working with human data, by analysing the content of academic guidelines, and built the conceptual framework ourselves. This framework was then used as input in the interview process. This allowed for the conceptual framework to be modified and refined, thereby resulting in a suitable and fitting framework to examine register researchers approaches to research ethics. The inductive framework seems like a valuable tool to explore research ethics in register research even further.

From the findings, it becomes evident that conducting register research comes together with ethical difficulties, and developments register researchers do not always approve of. There are constraints to register research that make it impossible to for example, withdrawal from the research, or to choose which data are shared with researchers. Register researchers do everything in their power to conduct research in an ethically sound way, but in some situations, there is a mismatch between the daily practice and the values they find important.

With big data research becoming increasingly more common and important in both public and private research environments, the gap between researchers' values and practice might increase. Register research could become the new 'Humphreys' research, the study where a sociologist observed men looking for intimate contacts in public restrooms, without informing them about the research or obtaining consent (Giddens, 2006; Henslin, 2012). Wanting to know more about these men, Humphreys looked up their licence plate in records, and went on to gather more information about them by visiting them at home. Instead, researchers conducting register research do not have to go into the field, or actively look for ways to connect information; data about peoples most intimate moments already exists and is available for researchers to analyse, as long as they have a solid proposal. It could be that one day, we, as society, look back at this situation and think: well, that was not how you are supposed to work with human data.

This requires vigilance with regard to research ethics, even in research fields where there is never any contact between researchers and the people they study, so important. This current study explored how register researcher approach the research ethics of human data, and identified several ethics and practices that register researchers perceive to be of importance in their work. Future research could expand on these findings even more, for example by conducting an in-depth study focussed on perceptions and attitudes towards notions of privacy in register research. Another research avenue is related to motivations and perceptions about research ethics in human non-subject research. We identified several motivations researchers have to take research ethics into account, but the range of these motivations could be investigated even more. Furthermore, this study identified different understandings of these protocols, but only gave a glimpse of what ethics are considered of importance, what motivations researchers have to adhere to protocols, and in what situations this is considered to be difficult. Furthermore, future research could incorporate more ethics guidelines (for example (Centraal Bureau voor de Statistiek, n.d.-b; MarktOnderzoekAssociatie et al., 2010; Sauerwein & Linnemann, 2002; Vereniging van Universiteiten, 2005), and broaden the research context to include more organizations, or even study the topic in an international setting. The findings opened up new research avenues allowing us to gain a deeper and broader understanding of how register researchers approach and practice research ethics.

5.4 Conclusion

Research ethics in register-based population studies are approached through practices of consent, purpose limitation, separation of functions, de-identification, assessment of risks and benefits, and dissemination of findings. Competence helps researchers form perceptions about what ethically sound register research entails. These findings reveal how register researchers working in the Dutch context approach the ethics of working with human data in a changing context, where information is increasingly available and enforcement on regulations is stricter. Future research can utilize the developed inductive framework to examine researchers' perceptions, attitudes and motivations towards research ethics even further. More information is needed on the distinction between values held by researchers and practices applied in register research. The findings of this study add to the current understanding of research ethics in theory and practice, and show how register researchers protect the interests of those included in register datasets.

REFERENCES

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct. Retrieved from http://www.apa.org/ethics/code/

Autoriteit Persoonsgegevens. (n.d.). Onderzoek en handhaving [Research and enforcement]. Retrieved August 1, 2017, from https://autoriteitpersoonsgegevens.nl/nl/over-de-autoriteit-persoonsgegevens/taken-en-bevoegdheden/onderzoek-en-handhaving

Autoriteit Persoonsgegevens. (2015). CBP krijgt boetebevoegdheid en wordt Autoriteit Persoonsgegevens [CBP becomes authorized to charge fines and becomes Data Protection Authority]. Retrieved from https://autoriteitpersoonsgegevens.nl/nl/nieuws/cbp-krijgt-boetebevoegdheid-en-wordt-autoriteit-persoonsgegevens

Babbie, E. (2010). *The practice of social research* (12th ed.). Belmont, CA: Wadsworth.

Bakker, B. F. M. (2009). Trek alle registers open! [Open up the registers!]. Amsterdam. Retrieved from https://research.vu.nl/files/2550298/Oratie Bakker.pdf

Bazeley, P. (2013). *Qualitative data analysis: Practical strategies* (1st ed.). London: Sage.

Brothers, K. B., & Wright Clayton, E. (2010). Human non-subjects research: Privacy and compliance. *The American Journal of Bioethics*, *10*(9), 15–17.

Centraal Bureau voor de Statistiek. (n.d.-a). Microdata: Zelf onderzoek doen [Microdata: Conducting own research]. Retrieved from https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen

Centraal Bureau voor de Statistiek. (n.d.-b). *Richtlijnen voor On Site/ Remote Access -output [Guidelines for On Site/ Remote Access - output]*. Retrieved from https://www.cbs.nl/-/media/8db64c6ff1ee41d5bce09e6b8628cff9.ashx

Dawson, L. M. (1997). Ethical differences between men and women in the sales profession. *Journal of Business Ethics*, *16*(11), 1143–1152.

Dolgin, E. (2014). Human-subjects research: The ethics squad. *Nature*, *514*(7532), 418–420.

Dowling, R. (2009). Ethical issues in research. In *International Encyclopedia of Human Geography* (pp. 595–600). Elsevier.

Drotar, D. (2011). Contemporary directions in research ethics in pediatric psychology. *Journal of Pediaric Psychology*, *36*(10), 1063–1070.

Eaton, A. A., & Visser, P. S. (2008). Attitude importance: Understanding the causes and consequences of passionately held views. *Social and Personality Psychology Compass*, *2*(4), 1719–1736.

Erwin, P. M. (2011). Corporate codes of conduct: The effects of code content and quality on ethical performance. *Journal of Business Ethics*, *99*(4), 535–548.

European Science Foundation and ALLEA. (2011). *The European code of conduct for research integrity*. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf

Friese, S. (2012). *Qualitative data analysis with ATLAS.ti* (1st ed.). London: Sage.

Giddens, A. (2006). *Sociology* (5th ed.). Cambridge, United Kingdom: Polity.

Guillemin, M., & Gillam, L. (2004). Ethics, reflexivity, and “ethically important moments” in research. *Qualitative Inquiry*, *10*(2), 261–280. Retrieved from file:///Users/beguemadeleine/Documents/Guillemin.pdf

Hedgecoe, A. (2008). Research ethics review and the sociological research relationship. *Sociology*, *42*(5), 873–886.

Hennink, M., Hutter, I., & Bailey, A. (2011). *Qualitative research methods* (1st ed.). London: Sage.

Henslin, J. M. (2012). *Sociology - A down-to-earth approach* (11th ed.). New York City, USA: Pearson.

Hsieh, H.-F., & Shannon, S. E. (2005). Three appraoches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288.

Hunt, A. (1997). Are women more ethical than men? *Baylor Business Review*, *15*(1), 7.

Israel, M. (2015). *Research ethics and integrity for social scientists* (2nd.). Thousand Oaks, CA: Sage.

Koller, S. H. (2008). Ethics in research with human beings: Some issues about psychology. *Ciência and Saúde Coletiva*, *13*(2), 399–406.

Komić, D., Marušić, S. L., & Marušić, A. (2015). Research integrity and research ethics in professional codes of ethics: Survey of terminology used by professional organizations across research disciplines. *PLoS ONE*, *10*(7), 1–13.

Koninklijke Nederlandse Akademie van Wetenschappen. (2012). *Zorgvuldig en integer omgaan met wetenschappelijke onderzoeksgegevens [Responsible research data management and the prevention of scientific misconduct]*. Retrieved from https://www.knaw.nl/nl/actueel/publicaties/zorgvuldig-en-integer-omgaan-met-wetenschappelijke-onderzoeksgegevens

Lehnert, K., Craft, J., Singh, N., & Park, Y.-H. (2016). The human experience of ethics: A review of a decade of qualitative ethical decision-making research. *Business Ethics: A European Review*, *25*(4), 498–537.

MarktOnderzoekAssociatie, Vereniging voor Beleidsonderzoek, & Vereniging voor Statistiek en Onderzoek. (2010). *Gedragscode voor onderzoek en statistiek [Code of conduct for research and statistics]*. Retrieved from http://www.vsonet.nl/images/bestanden/Gedragscode.pdf

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Thousand Oaks, CA: Sage.

O'Dwyer, B., & Madden, G. (2006). Ethical codes of conduct in Irish companies: A survey of code content and enforcement procedures. *Journal of Business Ethics*, *63*(3), 217–236.

O'Hanlon, G., Szép, K., & Smrekar, T. (2015). *Peer review report on compliance with the Code of Practice and the coordination role of the National Statistical Institute: Netherlands*. Retrieved from http://ec.europa.eu/eurostat/documents/64157/4372828/2015-NL-Report/05d4a3fd-8013-4fa5-bd0e-b38c3f54702b

OECD Global Science Forum. (2007). *Best practices for ensuring scientific integrity and preventing misconduct*. Retrieved from https://www.oecd.org/sti/sci-tech/40188303.pdf

Perryer, C., & Scott-Ladd, B. (2014). Deceit, misuse and favours: Understanding and measuring attitudes to ethics. *Journal of Business Ethics*, *121*(1), 123–134.

Pickens, J. (2005). Attitudes and perceptions. In N. Borkowski (Ed.), *Organizational Behavior in Health Care* (pp. 43–75). Sudbury, MA: Jones and Bartlett Publishers.

Pimple, K. D. (2002). Six domains of research ethics: A heuristic framework for the responsible conduct of research. *Science and Engineering Ethics*, *8*(2), 191–205.

Resnik, D. B., Neal, T., Raymond, A., & Kissling, G. E. (2015). Misconduct definitions adopted by U.S.

research institutions. *Accountability in Research*, *22*(1), 14–21.

Resnik, D. B., Rasmussen, L. M., & Kissling, G. E. (2015). An international study of research misconduct policies. *Accountability in Research*, *22*(5), 249–66. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4449617&tool=pmcentrez&rendertype=abstract

Richardson, S., & McMullan, M. (2007). Research ethics in the UK: What can sociology learn from health. *Sociology*, *41*(6), 1115–1132.

Rijksoverheid. (n.d.). Wie mag mijn gegevens in de Basisregistratie Personen (BRP) inzien? [Who is authorized to see my municipal personal record data?]. Retrieved March 6, 2017, from https://www.rijksoverheid.nl/onderwerpen/persoonsgegevens/vraag-en-antwoord/wie-heeft-recht-op-inzage-in-mijn-gegevens-in-de-gemeentelijke-basisadministratie-gba

Rothstein, M. A. (2010). Is deidentification sufficient to protect health privacy in research? *Americal Journal of Bioethics*, *10*(9), 3–11.

Saldaña, J. (2011). *Fundamentals of qualitative research*. New York, NY: Oxford University Press.

Sauerwein, L. B., & Linnemann, J. J. (2002). *Handleiding voor verwerkers van persoonsgegevens [Guideline for processors of personal data]*. Retrieved from https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/handleiding-wet-bescherming-persoonsgegevens.pdf

Shuster, E. (1997). Fifty years later: The significance of the Nuremberg Code. *The New England Journal of Medicine*, *337*(1), 1436–1440.

Singhapakdi, A., Kraft, K. L., Vitell, S. J., & Rallapalli, K. C. (1995). The perceived importance of ethics and social responsibility on organisational effectiveness. *Journal of the Academy of Marketing Science*, *23*(1), 49–56.

Sociaal-Wetenschappelijke Raad. (2006). *De toegankelijkheid van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek [The accessability of data from public and semi-public registers for scientific research]*. Retrieved from https://www.knaw.nl/nl/actueel/publicaties/de-toegankelijkheid-van-gegevens-uit-publieke-en-semi-publieke-administraties-voor-wetenschappelijk-onderzoek/@@download/pdf_file/90000128.pdf

Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review*, *154*(3), 477–564.

Stark, L. (2007). Review of: Voker Roelcke and Giovanni Maijo (Eds.). Twentieth centruty ethics of human subjects research: Historical perspectives on values, practices and regulations. *Journal of the History of the Behavioral Sciences*, *43*(2), 227–228.

Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity. (2010). *The Singapore statement on research integrity*. Retrieved from http://www.singaporestatement.org/statement.html

Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity. (2013). *The Montreal statement on research integrity*. Retrieved from http://www.researchintegrity.org/Statements/Montreal Statement English.pdf

Stein, L., & Sinha, N. (2002). New global media and communication policy: The role of the state in the twenty-first century. In L. Lievrouw & S. Livingstone (Eds.), *Handbook of New Media: Social Shaping and COnsequences of ICTs* (pp. 410–431). London: Sage.

The Netherlands. (n.d.). Wat zijn mijn rechten en plichten in de bijstand? [What are my rights and obligations when on welfare?]. Retrieved August 7, 2017, from https://www.rijksoverheid.nl/onderwerpen/bijstand/vraag-en-antwoord/wat-zijn-mijn-rechten-en-plichten-in-de-bijstand

The Netherlands. Wet basisregistratie personen [Municipal Personal Records Database act] (2015). Retrieved from http://wetten.overheid.nl/BWBR0033715/2015-09-01#Hoofdstuk3

The Netherlands. Wet bescherming persoonsgegevens [Personal data protection act] (2017). Retrieved from http://wetten.overheid.nl/BWBR0011468/2017-07-01

The University of Utah. (n.d.). Non-Human Subject Research. Retrieved August 5, 2017, from https://irb.utah.edu/guidelines/human-subject-research.php

United States. (1949). *Nuremberg code. Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law* (Vol. 10). Washington, D.C. https://doi.org/10.1136/bmj.313.7070.1448

United States. (1978). *The Belmont report: Ethical principles and guidelines for the protection of human subjects*. Retrieved from https://videocast.nih.gov/pdf/ohrp_appendix_belmont_report_vol_2.pdf

United States. (2016). Code of federal regulations: Title 45, part 46.102, *1*.

University of Groningen. (2012). *University of Groningen regulations for the protection of academic integrity*. Retrieved from https://www.rug.nl/about-us/organization/rules-and-regulations/algemeen/gedragscodes-nederlandse-universiteiten/wetenschappelijke-integriteit-12-en.pdf

Valentine, G. (2003). Geography and ethics: In persuit of social justice - Ethics and emotions in geographies of health and disability research. *Progress in Human Geography*, *27*(3), 375–380.

Valentine, G. (2005). Geography and ethics: Moral geographies? Ethical commitment in research and teaching. *Progress in Human Geography*, *29*(4), 483–487.

Velasquez, M. ., & Rostankowski, C. (1985). *Ethics: Theory and practice* (1st ed.). Upper Saddle River, NJ: Prentice Hall.

Vereniging van Universiteiten. (2005). *Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek [Code of conduct for the use of personal data in scientific research]*. Retrieved from http://www.vsnu.nl/files/documenten/Domeinen/Accountability/Codes/Gedragscode persoonsgegevens.pdf

Vereniging van Universiteiten. (2014). The Netherlands code of conduct for academic practice: Principles of good academic teaching and research. Retrieved from http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_%282012%29.pdf

Willenborg, L., & Heerschap, N. (2010). *Koppelen [Matching]*. Den Haag/ Heerlen. Retrieved from https://www.cbs.nl/-/media/imported/onze diensten/methoden/gevalideerde methoden/throughput/documents/2010/18/2010-x37-13-pub.pdf?la=nl-nl

APPENDIX 1 CODE TREE RESULTING FROM CONTENT ANALYSIS OF CODES OF RESEARCH CONDUCT

This appendix displays the code tree that resulted from the content analysis of eight academic codes of research conduct. The first order of the tree displays the theme, the second order shows the subthemes, and with an example of a quotation included under the subtheme.

Competence

- Researcher Qualifications
    - “A scientific practitioner ensures that he maintains the level of expertise required to exercise his duties. He does not accept duties for which he lacks the necessary expertise. If necessary, he actively indicates the limits of his competence.” (Vereniging van Universiteiten, 2004:5)
- Adherence to protocols
    - “The collaboration as a whole should be in compliance with all laws, policies and regulations to which it is subject.” (Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2013:1).

Respect for persons

- Consent
    - “Human subject protocols should not be violated: this implies complying with the requirement of informed consent on the basis of adequate and appropriate information, and to voluntary agreement to participate [...].” (European Science Foundation and ALLEA, 2011:14).
- Privacy
    - “Deviations from desired [research] practices include [...] insufficient privacy protection […].” (European Science Foundation and ALLEA, 2011:12)

Beneficence

- Assessment of risks and benefits
    - “Accordingly, so-called risk/benefit assessment are concerned with the probabilities and magnitudes of possible harms and anticipated benefits. Many kinds of possible harms and benefits need be taken into account” (United States, 1978:15)

Justice

- Selection of subjects
    - “[...], the principle of justice gives rise to moral requirements that there be fair procedures and outcomes in the selection of research subjects. Justice is relevant to the selection of subjects of research at two levels: the social and the individual.” (United States, 1978:18)

APPENDIX 2 OVERVIEW CODES OF RESEARCH CONDUCT EMPLOYED FOR CONTENT ANALYSIS

Ethics practices included in codes of research conduct.

	Nuremberg Code (United States, 1949)	Belmont Report (United States, 1978)	Best Practices for Ensuring Scientific Integrity and Preventing Misconduct (OECD Global Science Forum, 2007)	The Singapore Statement on Research Integrity (Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2010)	The Montreal Statement on Research Integrity (Statement Drafting Committee at the 2nd and 3rd World Conference on Research Integrity, 2013)	The European Code of Conduct for Research Integrity (European Science Foundation and ALLEA, 2011)	The Netherlands Code of Conduct for Scientific Practice (Vereniging van Universiteiten, 2014)	University of Groningen Regulations for the Protection of Academic Integrity (University of Groningen, 2012)
Policies and regulations	.	.	X	X	X	X	.	.
Researcher qualifications	X	.	.	.	.	.	.	.
Consent	X	X	.	.	.	X	X	.
Assessment of risks and benefits	X	X	.	X	.	X	X	.
Selection of subjects	.	X	.	.	.	X	.	.
Privacy	.	.	X	.	.	X	X	.

APPENDIX 3 INVITATION TO PARTICIPATE IN SEMI-STRUCTURED IN-DEPTH INTERVIEW

Subject: Research ethics practices in register data research – Invitation to participate in study

Dear <Name>,

My name is Rosanne Spruijt, and I am a Population Studies master student studying at the University of Groningen. For my master thesis project, I explore how population researchers who study population register data approach the topic of research ethics, more specifically: the research ethics of working with human data. I would like to invite you, a professional who deals with these data on a daily basis, to have a discussion with me about the research ethics you apply in your work.

I plan on interviewing professionals working at several Dutch research organizations, and I would be very grateful if you would want to participate in my study. My goal is to hear a diverse range of stories about this topic, not to check your work, or look for misconduct. The interviews are expected to last approximately ***45 minutes****, and will be audio-recorded (and/or notes will be taken during the discussion). The information you provide will be treated confidential, which means personal names or any other information which would serve to identify you as an informant will not be included in any reports or presentations resulting from this project.*

If you can accept this invitation to participate, or have any questions or concerns about the study, please contact me per email (e.r.spruijt@student.rug.nl)! You can also contact my supervisor for this project, dr. Ajay Bailey, assistant professor at the Faculty of Spatial Sciences of the University of Groningen, per email (a.bailey@rug.nl) or phone (050-3633898) with any questions you might have.

Thank you for your consideration,

Kind regards,
Rosanne Spruijt

APPENDIX 4 INTERVIEW GUIDE USED IN SEMI-STRUCTURED IN-DEPTH INTERVIEWS

Introduction

Thank you so much for having me, and letting me ask you some questions. Like I wrote in the email, I am conducting my master thesis research at the university of Groningen, and I am interested in hearing about how you approach research ethics in your work as a population researcher. I hope that you want to share your thoughts and opinions about the research *practices* that come together with doing research on human data. So: I would like to know which research ethics practices you apply; which practices you find especially important; and why. My goals is to, in the end, understand how you deal with research ethics in your work. I would also like to get insight in the role of 'codes of conduct' in your work, to see whether these codes are helpful in making decisions in population register research. So that's why I would like to ask you some questions. Would an interview of about 45 minutes work for you?

Confidentiality

I'd like to hear your thoughts and opinions about the topic, so there are no right or wrong answers to give in this interview. Everything you tell me will be treated confidential, which means that I will not share the things you tell me in such a way that they can be traced back to you. I will use the information you tell me in publications and presentations, but I will leave out the parts that can be used to identify you.

I will be taking some notes during the discussion, and if you're ok with it, I would like to record the audio of our conversation so I can listen to it later on. I would not share the recordings with anyone, except for my supervisor, and I will keep the recordings in a password protected file on my computer for up to two years.

I told you a bit about the purpose of this interview, and we discussed how I will treat the information you provide. As a participant in this study, you have the right to decline participation, decline to answer any particular question, ask for the recorder to be turned off, or to end the discussion at any time. You have the right to fully withdraw from this project, or ask for erasure of any materials you do not wish to be used in any reports of this project, up until two weeks after participating in the discussion. You are, of course, free to ask any question you might have about the research project before, during, or after your participation! Do you have any questions for me about this, or about something else?

Consent

Ok, if you would want to sign here to confirm your consent? This is the email I invited you with, which also includes the information about the project and confidentiality.

➔ Start audio recording and the interview.

Opening question

1. Can you tell me a little bit about your work?

Probe: How long working for “organization” // Education // Topic of interest // Typical workday

2. I invited you for this interview because you work with population register data. Can you tell me about the process which allows you to get access to these data?

Probe: Lay over a research proposal // Online or local access

Questions about respect for persons/autonomy

3. What do you think ‘consent’ is about?
4. Can you tell me about the role of consent in your work?

Probe: Opinion // Process // Informed? // Permission? // Withdraw

5. What do you think privacy means?
6. Could you tell me something about the role of privacy in your own work?

Probe: Opinion // Process of de-identification // Security measures

Questions about beneficence

7. Can you tell me a little bit about the risks and benefits of your work?

Probe: Risks // Benefits // Process of assessment // When applicable // Opinion

Questions about justice

8. Can you tell me something about the process of data selection in your work, so: how is determined whose data gets included in the research?

Probe: Opinion // // Relevance of sensitivity in data selection// When and Why?

Questions about competence

10. Can you tell me something about the role of codes of conduct in your work?

Probe: Code “organization” // Know any other codes // Suitable // What helps // What is missing

11. Can you tell me a little bit about maintaining or improving your level of expertise as a population register researcher?

Probe: Education or training // congresses // consult colleagues about research ethics

12. How do you decide which practices to apply in your work?

Probe: Formal education // Experience // Consult/discuss with colleagues // Guidelines // Other

Closing questions

13. Are there research practices related to working with population register data that you would like to see implemented or improved in your field?

14. Have your thoughts about which practices you find important changed since you started your career? How?

That was the last question I have for you. Is there anything we have not yet talked about, that you would like to discuss?

➔ Closing discussion, thanking participant

APPENDIX 5 CODE TREE RESULTING FROM THEMATIC ANALYSIS OF INTERVIEW DATA

Code family: Perceptions/attitudes about research ethics applied in own work

- Purpose limitation ("Doelbinding")
- Separation of functions ("Functiescheiding")
- De-identification ("Herleidbaarheid voorkomen")
- Changing context
    - Change in protection and access to data
    - Enrich data through matching
    - Enforcing protocols

Code family: Perceptions/attitudes about human data research ethics

- Perception informed consent
    - Interested in macro, not micro
    - Consent responsibilities
- Perception privacy
    - Identifiability/herleidbaarheid
    - Confidentiality
    - Privacy-autonomy
    - Sensitivity of data as a scale
- Perception assessment of risks and benefits
    - Responsibility in assessment
    - Perceived benefits
    - Perceived risks
- Perception about distribution of risks and benefits
    - Responsibility in distribution of risks and benefits
    - Research question leads data selection/analysis/dissemination
    - Perceived benefits
    - Perceived risks

Code family: Competence

- Expertise
    - Learning in work environment
    - Self-study
    - Learning: institutional
- Protocols
    - Attitude protocols
    - Internalising contents

- Common sense
- Showcase mechanism
  - Credibility
- Enforcing protocols
- Perceived constraints
- Mentioned guidelines
  - Academic
  - VSO
  - Work instructions (“werkafspraken”)/ Individual arrangements
  - CBS
  - Law