Consent to link health data: older adults in New Zealand

A thesis presented in partial fulfilment of the requirements

for the degree of

Master of Arts

In

Psychology

At Massey University, Manawatū,

New Zealand.

Laura Jessie Crawford

2017

Abstract

With administrative data increasingly being recorded electronically, data linkage has become a popular method of research. It involves the linkage of two data sets – survey and administrative data, in the current study – to create a wider and more varied data set, with which a greater number of research questions can be examined. Seeking consent from participants to link their data is an ethical and legal requirement. However, consent seeking may create systematic bias as likelihood of consenting may be associated with a variety of health and socioeconomic variables. Variables associated with consent were examined for linkage between the Health, Work Retirement longitudinal study and Ministry of Health data sets in New Zealand. Unlike previous studies of this type, participants were older adults. Binary logistic regression revealed that Māori ethnicity (OR 0.68), Diabetes (OR 0.66), and participating for more survey waves (OR 1.88) were significantly associated with consent. The model explained 7 to 10% of the variance in consent, suggesting that older adults are not greatly influenced by these variables. Implications for research and policy are discussed.

Acknowledgements

Contents

Consent to link health data: older adults in New Zealand

A quick introduction to data linkage

Often data about the same individual will be spread across a variety of different data sets. For example, data collected by a hospital while offering treatment (administrative data), responses to surveys sent out by universities (survey data), and information collected as part of a census. Often these pieces of information remain separate. However, it can often be beneficial to bring these disparate sources together for research purposes. This is data linkage. It involves taking one or more sets of data that relate to the same people, and identifying which records relate to the same individuals to create a larger data set (Bohensky et al., 2010). The researcher can then, for example, learn about how location (recorded in the census) impacts on depression (recorded in survey and administrative health data; e.g. Maguire & O'Reilly, 2015). Data linkage is an increasingly popular research method because it enables us to shed light on crucial, unanswered questions in health research.

Linking survey data to administrative data sets enables us to examine patterns that either set alone cannot. For instance, administrative data reliably captures things like diagnosis, hospital attendance, and prescription use. These are collected while providing a service and must be accurate for the service to be effective. However, administrative data doesn't usually capture some individual level characteristics, risk factors, or attitudes, which are captured in some survey data (Madans & Cohen, 2005). Linking these two elements together opens a range of new possibilities for study, such as the association between attitudes towards medical care and prescription use, or how perceived quality of life influences the use of primary healthcare services. These types of analyses become increasingly crucial to healthcare policy and practice as limited resources are stretched to cater to an ageing population (Ha, Hendrie, & Moorin, 2014).

While data linkage is powerful, there are issues that can prevent linkage from producing useful, representative data sets. A key factor that may reduce the adoption of data linkage for research purposes is the common requirement that consent be sought from individuals whose data is being linked. The process of gaining informed consent can reduce the resulting sample size under study as not all participants will consent (Sakshaug, Couper, Ofstedal, & Weir, 2016). Perhaps more importantly, a reduction in sample size due to the consent process can result in systematic differences between the population under study and the data linked sample. This is called consent bias.

This bias, that may result from the process of gaining consent for data linkage between large datasets (e.g. survey and administrative datasets, two large administrative datasets), should be of special concern for researchers in the health field. Health status, and variables representing health status, is one variable that has been consistently related to consent to linkage; participants with poorer health tend to consent more readily (Dunn, Jordan, Lacey, Shapley, & Jinks, 2004). This in turn can impact the results of research using the data that has been linked (Ioannidis, 2013). Given the follow-on impact of such research on healthcare policymaking and practice, biased linked data has the capacity to negatively impact the health of the populations it is supposed to shed light on.

This thesis examines an instance of data linkage between longitudinal health and ageing survey data from the Health, Work, and Retirement study (HWR; Dulin, Stephens, Alpass, Hill, & Stevenson, 2011; Towers, Stevenson, Breheny, & Allen, 2017) and administrative health data held by the Ministry of Health. It will highlight any limitations of the HWR linkage and areas of concern for researchers using the data. Additionally, consent bias has not been examined in an ageing population and so this analysis will add to the existing literature on consent bias.

Literature Review: Consent to Data Linkage in Health and Ageing Research

Robust, useful healthcare research is vital to understanding population health status, for improving healthcare services, and ensuring that public health issues are identified and addressed effectively (OECD, 2017). Large, international health surveys contribute greatly to our understanding of public health issues. They highlight trends and issues in health across countries.

WHO's World Health Survey (World Health Organization, 2017), implemented in 2002-2004, is a key example of valuable surveys. It covers 70 countries and 300,000 individuals, selected to be representative of the entire world. The survey covers both individual and household data. Household data includes a household roster, health insurance coverage, health expenditures, and indicators of permanent income or wealth.

Individual level data includes:

- sociodemographic information
- health state descriptions
- health state valuation
- risk factors
- chronic conditions

- mortality
- health care utilisation
- health systems responsiveness
- social capital

Data from the survey are made available on request to researchers and forms the basis for hundreds of studies since 2005 (World Health Organization, 2017). The survey data is a unique and invaluable tool for examining similarities and differences in health trends and issues between countries, as well as tracking health trends within countries.

Recent studies using the World Health Survey data have found, for example, that depression and inflammatory arthritis are associated in both high- and low-income countries, indicating that this is a global phenomenon (Apfelbacher, Brandstetter, Herr, Ehrenstein, & Loerbroks, 2017). This suggests a greater link between the two issues and may spur further research on the association. Another study found that Gross Domestic Product (GDP) is associated with multi-morbidity prevalence across countries (Afshar, Roderick, Kowal, Dimitrov, & Hill, 2015), highlighting the inequalities in ageing experience between countries.

In New Zealand, the Ministry of Health conducts the New Zealand Health Survey (NZHS). This survey exists to support the development of health services, policy, and strategy. It assesses various health domains of concern to the New Zealand public and policy makers. These include:

- self-rated general health

- various health-related behaviours including tobacco and alcohol use, nutrition, physical activity, body size, and physical punishment for children

- various health conditions including cardiovascular health, mental health conditions, diabetes, and chronic pain

- access to healthcare, such as primary care use and unmet healthcare need

- oral health

The 2015/2016 NZHS report highlighted several trends in New Zealand health behaviors and healthcare use that are of concern to researchers and policy makers (Ministry of Health, 2016). The report found that the rate of smoking has decreased, but hazardous patterns of alcohol consumption have increased. One in five New Zealand adults was found to have a hazardous drinking pattern. They also found that rates of unmet need for medical care are increasing for both adults and children, and that Māori and Pacific Islander groups have an unmet health need 2 to2.5 times greater than New Zealand Europeans (Ministry of Health, 2016). These trends are of concern for practitioners, policy makers, and the New Zealand public.

Surveys like these identify specific trends and attitudes in health but may lack reliable, concrete medical data critical for understanding the outcomes of these trends on the health sector (e.g. hospitalisation use, prescriptions, mortality) and other areas. While this kind of survey data is incredibly valuable, it often lacks the depth necessary to measure impact of trends on the health system itself and the costs associated with maintaining or improving health services (e.g. Carter, Shaw, Hayward, & Blakely, 2010).

Increasing focus on data linkage in health research

Data linkage makes use of existing datasets, often collected for other purposes such as healthcare provision, by joining them to other sets of data to answer new questions. More specifically data linkage is "a process of pairing records from two files and trying to select the pairs that belong to the same entity" (Winglee, Valliant, & Scheuren, 2005 quoted in Bohensky et al., 2010). The term 'record linkage' is sometimes used synonymously with data linkage (Winkler, 1999, para. 1) although others use the term more specifically to refer to data about a person or entity (e.g. Black & Roos, 2005; Dunn, 1946; Newcombe 1988). In the context of healthcare, the units of observation are most frequently individuals or families, and so for the purposes of this thesis they should be considered synonymous.

Data linkage methods were initially established in the late 1960s (Fellegi & Sunter, 1969) and has increased in popularity since. Felligi and Sunter (1969) suggested that the increasing popularity of data linkage could be accounted for by increased accumulation of administrative data, increasing data processing capability, and awareness of research uses, though several other explanations exist including cost effectiveness and data quality (Glasson & Hussain, 2008; Martin-Sanchez & Verspoor, 2014). The power of data linkage to answer questions about healthcare outcomes remains a key aspect of its continued popularity and expansion.

Large linkage studies and databases are now maintained across the world, in countries such as Canada (Boyd et al., 2014), the United States (Melton, 1996), The United Kingdom, Denmark, Sweden, Norway, the Netherlands, Italy (Charlton et al., 2017), Australia (Holman et al., 2008), and New Zealand (Statistics New Zealand, 2012). Most of these data linkages contain primarily health data. The Western Australia Data Linkage System (WALDS) contains 30 health data collections for the population of Western Australia (Holman et al., 2008). It contains electoral, hospital, death, cancer, mental health and other records which support a wide range of research topics.

Health and healthcare information is increasingly recorded electronically (Stewart & Davis, 2016). This is true of administrative healthcare data, register and monitoring data, and data from related areas such as housing, which can help inform our understanding of health. Administrative healthcare datasets include things such as the number of people accessing a service, service use dates, diagnoses, mental health events, length of hospital stays, and medication prescriptions. Monitoring data can include things like registers for cancer and rare diseases (Bradley, Penberthy, Devers, & Holden, 2010). Without this trend towards electronic healthcare recording, many data linkage applications would not be possible.

These administrative datasets provide opportunities for researchers to answer questions about healthcare and how our healthcare system is performing, such as the current rates and length of hospitalisation. Including administrative datasets alongside ongoing health survey data significantly expands a researcher's capabilities to explore dimensions of health. For example, linkage can enable investigation into psychosocial or socioeconomic trends that result in increased hospitalisation rates or mortality outcomes.

Linked health information is required to answer the increasingly complex questions that arise in healthcare (Chamberlayne et al., 1998). Researchers can integrate more variables into their models, allowing them to develop more complete pictures of health and wellbeing, and facilitating an assessment of the influence of health and social factors on healthcare utilisation. Large data sets, such as longitudinal, linked, and population-level datasets, are necessary for reliable research about

outcomes for minority groups (Black & Roos, 2005; Glasson & Hussain, 2008). Data linkage also provides opportunities for researchers to use data outside what is generally considered health data. By linking with data from outside the healthcare system, such as geographic data, researchers can incorporate wider contextual variables to better understand health determinants and solutions (Regidor, 2004; Roos et al., 2008).

For example, Borschmann et al. (2016) identified a gap in knowledge about self-harm in recently released prisoners. Despite known links between prison populations and suicide as well as self-harm and suicide, no research had investigated the incidence of self-harm among prisoners. To understand the incidence of self-harm in recently released prisoners, Borchmann and colleagues linked survey data, correctional health records, and emergency department health records. They discovered that the rate of self-harm at follow-up was at 10 times what was usually reported in the general population (Borschmann et al., 2016).

Another example of effective data linkage use is Pickrell and colleagues' investigation of the reason for the link between epilepsy and deprivation (Pickrell et al., 2015). While the association between epilepsy diagnosis and deprivation had been well-established, researchers could not identify the cause of the link. To do this, Pickrell et al. (2015) linked electronic primary health records with deprivation scores. The authors also followed a cohort of 582 participants for 10 years to determine whether that decile changed over time, and found that it did not. This showed that deprivation influences the incidence of epilepsy rather than epilepsy causing sufferers to become more deprived over time. The authors suggest that this may be the first time such an analysis has been done, despite its implications for understanding how epilepsy develops and how epilepsy risk factors can be addressed (Pickrell et al., 2015).

In addition to increasing the scope of what can be investigated, data linkage methods have inherent efficiencies that make it an attractive choice for researchers. Linked data sets typically contain a mix of information created by researchers, such as longitudinal survey data, and data collected and re-purposed from pre-existing sources (Black & Roos, 2005). This is a deceptively powerful quality. It means that linked datasets can include a complex set of disparate datasets, longitudinal trends, and administrative data to better understand hard outcomes. Linked databases can also be accessed many times to answer different questions. These features minimise the need for unnecessary data collection (Bradley et al., 2010) and reduce the cost of obtaining data. This makes linkage economically efficient for universities and governments (Lipowski & Bigelow, 1996).

Features of data linkage also create benefits for research participants (Bohensky et al., 2010; Huang, Shih, Chang, & Chou, 2007). The pre-existing administrative data is re-used rather than researchers

collecting that data again directly from participants. This minimises participant burden. Not only can participant burden result in lower rates of research participation but it can also create difficulties for participants and lead to mistrust in research. Participant burden is particularly problematic for minority populations, who may have previously been over-researched. This can harm groups with healthcare issues where further research is needed (Pagano-Therrien, 2013; Sharp & Frankel, 1983).

Additionally, linkage presents a clear benefit to government and policy makers. To make robust decisions, government and policymakers must have access to high-quality data from which reliable conclusions can be drawn about various groups and the broader population. To this end linked longitudinal datasets can facilitate the analysis of health and hospitalisation trends over time or by specific factor, such as chronic conditions, with considerable efficiency. Large datasets facilitated by linkage enable clear conclusions even for small groups, such as individuals with less common conditions. Sound conclusions from such data can support resource allocation, decision making, and policy development.

Health and ageing research using data linkage

As with other areas of health research, data linkage methods are becoming increasingly common in health and ageing research. Linked data allows researchers to answer otherwise challenging or impossible questions about health outcomes in older adult and ageing populations (Black, 1995). This section discusses the prevalence and importance of data linkage in health and ageing research. It showcases linkage studies and highlights how the valuable knowledge they generate could only be gained through linkage techniques.

As our population ages, age-related healthcare utilisation and potential healthcare burden becomes an increasingly urgent policy issue. This is particularly true of the care and management associated with chronic illnesses, the development of which increases with age (World Health Organization, 2015). While ageing is associated with greater health burden, use of health services does not always straightforwardly increase with age across all communities. For example, access to services can be limited in poorer communities (World Health Organization, 2015). Addressing issues of healthcare access and investment for older adults therefore requires both comprehensive data to shed light on access, usage, and resulting impact on health, and an understanding of the sociodemographic characteristics of the individuals likely to seek health services.

Healthcare records provide only part of the story because they only capture events that take place within the healthcare system (e.g. prescriptions, test results, hospital readmission and emergency room visits). They do not routinely include social or general daily wellness information. Data linkages

provide opportunities for researchers to answer questions about wellbeing in older age, health behaviours, and healthcare usage that health records alone cannot.

The following literature review highlights instances of data linkage use in older adult and ageing healthcare. It shows how the questions asked by the authors could only have been answered using data linkage techniques, ultimately highlighting the importance of linkage as a research method.

Linked data can reveal reasons for health service use. The Netherlands has seen an increase in hospitalisation rates among the ageing population following several policy and operational changes that increased access to hospital care (Galenkamp, Deeg, de Jongh, Kardaun, & Huisman, 2016). The changes included certain treatments being increasingly performed during day admissions, relaxed budgetary constraints, and a new funding initiative that incentivised hospitals to minimise lengths of stay (Galenkamp et al., 2016). Researchers had been unable to identify if the increased hospitalisation rates were due to increased access, increasing disease burden among the ageing population, or both.

To address this question, individual level data was required. Galenkamp, Deeg, de Jongh, Kardaun, and Huisman (2016) linked morbidity data from the Longitudinal Ageing Study Amsterdam with the Dutch Hospital Discharge Register. This allowed them to determine what kinds of hospital admissions were increasing and what portion could be attributed to increased disease burden. They found that only increases in acute overnight admissions were accounted for by worsening health in the ageing population (Galenkamp et al., 2016). This research more clearly defined the issue for policy makers; although access contributed to the increased hospitalisation, there was also a genuinely increasing need for overnight hospital admissions in the Netherlands.

Other studies have also demonstrated the utility of linked data sets for examining health service use across a variety of settings. An Australian linkage study allowed the authors to demonstrate how hospital use in the last 12 months of life differed greatly depending on the cause of death (Harris, Dolja-Gore, Kendig, & Byles, 2016). Another Australian study demonstrated location-based inequality in access to health care for those who had fall-related injury (Sukumar, Harvey, Mitchell, & Close, 2016).

Mortality data linkage enables study of associations. To determine whether a given condition or experience is associated with mortality, the two sets of data must be linked. This may require linking two sets of administrative data, such as where the health condition of interest is recorded in health registries. Alternatively, it may require linking administrative data with survey data, such as where

the researcher is interested in understanding an experience or perception and its association with ageing.

To investigate the impact of perceived stress on mortality in a Taiwanese population, Vasunilashorn, Glei, Weinstein, & Goldman (2013) linked data from the Survey of Health and Living Status of the Elderly and Near-Elderly with death certificate files held by the Taiwan Department of Health. The authors created several models covering an 11-year follow up period. In the first model they found an increase in perceived stress by one standard deviation represented a 19% mortality risk increase.

However, this relationship disappeared when medical conditions, mobility limitations, and depressive symptoms were taken into account (Vasunilashorn et al., 2013). This linkage study could expand on previous findings by using a survey with a broader measure of perceived stress and controlling for health, social support and behavioural variables. The authors sum up their findings by highlighting that they fit into the established model that indicates health may mediate the relationship between perceived stress and mortality.

Other studies have also demonstrated the utility of linked data sets for understanding health associations with mortality. Researchers in Sweden used linked data to establish a link between all types of atrial fibrillation and increased risk of all-cause mortality and ischaemic stroke (Batra et al., 2016). Research revealing associations with mortality can inform best practice decisions in healthcare.

Linked data can show the impact of chronic conditions. As the population ages, the proportion of individuals who have survived cancer will increase. It is unclear how much of a health cost this group represents, making it important to understand quality of life in older adults who have survived cancer. Examining quality of life in cancer survivors requires two sets of data; a quality of life measure and a record of cancer incidences.

Kent et al. (2015) in the United States (US) linked survey data measuring quality of life with the Surveillance, Epidemiology and End Results (SEER) national cancer registry. Linking these data sources allowed the researchers to investigate quality of life among older adults who had survived less common forms of cancer such as kidney, bladder, pancreatic, upper gastrointestinal cancer, uterine, cervical, and thyroid cancers. They compared physical and mental health component scores on the SF-12 for cancer survivors and people who had not had cancer, and found the biggest differences in quality of life for patients with pancreatic cancer and multiple myeloma. By linking the SEER database, the authors were able to present quality of life differences for a wide variety of cancer types.

As with other health issues, heart-related health issues present a risk of considerably increasing costs as the population ages. To examine the healthcare costs associated with ageing among those with cardiovascular disease, Ha, Hendrie, and Moorin (2014) used the Western Australian Data Linkage System (WADLS). They were able to identify a sample of 25,126 'episodes of care' for people with cardiovascular disease from 1 July 1993 to 30 June 2004. Using data from the Hospital Morbidity Data System and assigning associated costs to each episode of care, the authors examined the impact of ageing on hospitalisation costs. The availability of the Hospital Morbidity Data System data created a large and detailed sample for the authors to examine, enabling them to account for population growth, increases in costs, different types of cardiovascular disease, and gender as well as age.

There was a complex relationship between population ageing and healthcare costs. Overall, the contribution of ageing in those with cardiovascular disease to increases in the cost of hospitalisation was 22.9% for men and 22.7% for women. Over half of this increase was accounted for by the increase in average cost per episode of care. For those with chronic conditions, the impact of ageing on hospitalisation costs was far greater than for those with acute coronary syndrome and stroke (Ha et al., 2014). This result illustrates the impact of the increase in chronic conditions in the ageing population and the need for healthcare systems to build capacity for dealing with chronic conditions (World Health Organization, 2015).

Other studies have also demonstrated the utility of linked data sets for examining the impact of chronic conditions. An Australian study found that participants with diabetes were 24% more likely to have a hospital admission for any reason in a year (Comino et al., 2015). Other studies in the US used linked data to suggest that clinical advice regarding cancer screening amongst older veterans is not being followed in the US, creating unnecessary health risks and increasing costs (Walter et al., 2009; Walter, Bertenthal, Lindquist, & Konety, 2006).

As these examples show, data linkage provides a means for researchers to investigate many health outcomes. Administrative information about service use, mortality, diagnosis, geography, and many other health outcomes can be linked to better understand how services are provided, used, and what opportunities there are to improve healthcare. This research, much of which would be impossible without linkage techniques, can inform best practice and support policy decision making. While these studies demonstrate the usefulness of data linkage for health research, the common requirement for informed consent from participants raises several issues for data quality. Asking consent from participants can bias the resulting sample, which in turn can easily bias research outcomes when data linkage is used in the health research field.

Consent bias can impact research quality

Data-linkage has provided researchers with a powerful avenue to explore complex unanswered health questions. However, one critical process underpinning the use of data-linkage as a research tool has the potential to undermine its utility. Before linkage takes place, researchers are commonly required to obtain consent from the individuals whose data is to be linked. This requirement presents a risk of consent bias and ultimately risks research quality. This can undermine the value of findings from data linkage research to researchers, practitioners, and policy makers.

To link two distinct datasets, they must each contain the same identifying information for each individual. This may include names, dates of birth, and identifying numbers. To obtain data with this information, or to provide this information to a third party to conduct a linkage, researchers must usually meet strict ethical requirements which include obtaining consent from their participants (Glasson & Hussain, 2008). In the consent gaining process participants are lost, sometimes in systematic ways, which can introduce bias.

The way in which a data linked study is set up not only determines who holds the data but can also present different issues for consent, data ownership, and confidentiality. Data linkages for research purposes can occur in a number of different study designs. Bradley et al. (2010) categorise these according to the data-controlling parties, or what they call 'major agents':

1. Individual investigators,

2. Government-sponsored linked databases, and

3. Public-private partnerships that facilitate the use and linkage of data owned by private organisations (p. 1469).

The design and data-controlling party influences how the data linkage is conducted. Things that may differ between designs include how long the data can be held, the size of the dataset, and whether participants are asked consent. For example, data linkages undertaken by public partnerships will often make data available to researchers from both institutions. This can create confidentiality risks not present if the linkage is conducted by one institution.

It is possible to obtain a population-level data set without asking for consent to linkage. This can occur when data linkage is undertaken by a government or government-sponsored organisation, using existing administrative datasets (e.g. records concerning healthcare utilisation, education, tax). In these cases, consent bias will not be an issue as consent does not need to be sought. However, for individual researchers and non-government organisations undertaking linkage (e.g. between

population surveys and administrative datasets) consent is almost always required by law and ethics committees, and so the analysis of consent bias is relevant to such designs (Dokholyan, Muhlbaier, & Falletta, 2009; Jutte, Roos, & Brownell, 2011).

Overall participants tend to consent to data linkage more often than they decline or make no response, although consent rates can vary considerably between studies. Meta-analyses and large studies have shown consent rates to link health-related data vary from 36.6% to 97% but are commonly over 50% (da Silva et al., 2012; Kho, Duffett, Willison, Cook, & Brouwers, 2009). This research shows that consent rates to data linkage can fluctuate considerably. Regardless, even for those studies with high consent rates, seeking consent from participants inevitably reduces the size of the sample.

When participants choose to consent or not in systematic ways, this can result in a sample that is no longer representative of the population under study. Consent bias refers to the idea that in seeking consent for data linkage researchers could be left with biased samples, and therefore produce biased results. Buckley, Murphy, Byrne, and Glynn (2007) say that consent bias describes "the selection bias resulting from the loss of non-consenters to any cohort" (p.1116). For example, if younger people are more likely to consent and the study is examining heart disease, which younger people are less likely to experience, then that health condition will be under-represented in the consenting sample. That may then influence the study outcomes.

The concept of consent bias arose from research showing significant differences in rates of consent between different demographic groups. An examination of the literature found that only 5% of data linkage studies examined bias in their consenting samples despite evidence of considerable bias (Bohensky et al., 2010). However, biased samples do not necessarily result in biased research findings.

For a variable to bias the results of a study, the variable on which the consenting and non-consenting groups differ must be systematically associated with both the dependent and independent variables under study (Rothstein & Shoben, 2013). This means that consent bias may not be an issue for all research projects. However, for projects in which sub-samples do differ on a covariate that may be related to the desired outcomes, it is important for the researchers to identify this as an issue. Understanding what variables are associated with consent enables researchers to not only gauge the consent rates they can expect for a study but also determine whether consent bias is likely to influence their results, based on the variables under study.

Consent bias is a sub-set of response bias

Some authors have noted that participant non-response will inevitably impact the sample size more than non-consent (Dunn et al., 2004). This is because consent is usually synonymous with opt-in consent. Participants must respond proactively to give their consent and refusing consent is as simple as not responding. As a result, most participants who respond also consent and most non-consenters are also non-responders. Almost every study examined here follows this pattern in non-consent.

This relationship between non-response and non-consent means that consent bias can be seen as a subset of response bias. Few researchers analyse non-consent and non-response separately, instead viewing passive non-consent and active non-consent as identical. This is sufficient for the purposes of understanding bias and eliminates the issue of attempting analysis with tiny active non-consent groups.

Consent bias likely impacts health research quality

Consent bias is very likely to influence research in the health field. Overall health status and other measures of health have been repeatedly associated with consent, both generally and to data linkage specifically (e.g. Buckley, Murphy, Byrne, & Glynn, 2007; Carter et al., 2010; Dunn et al., 2003). As a result, the variable influencing consent is highly likely to be related to the variables under study. With the growth in data-linkage projects being undertaken in health research, it is critical that researchers acknowledge the potential of consent bias to influence their findings and explore which variables are central to such bias.

In demonstrating the effect of consent bias, some researchers have taken a step further than examining correlations with consent. They have examined the effect of biased samples on research outcomes. These types of analyses show exactly how the biasing effect of consent on the sample could impact on research outcomes for a specific topic or study.

Al-Shahi, Vousden, and Warlow (2005) examined how consent bias influenced results in a study of participants with brain arteriovenous malformation, an abnormal connection between arteries and veins. Of those asked to participate, 59% consented. The authors examined differences between participants who consented and those who didn't but did not find significant demographic differences between consenters and non-consenters. Al-Shahi, Vousden, and Warlow then took their analysis a step further to understand the impact of excluding non-consenters from the sample on the results of analysis. The authors found that when only consenters were analysed, initial presentation with intracranial haemorrhage was not associated with risk of subsequent

haemorrhage. However, when the analysis included non-consenters a significant association was found (Al-Shahi et al., 2005).

If conducted without information from participants who declined, this analysis would suggest that intracranial haemorrhage was not a risk for subsequent haemorrhage. This finding could then influence healthcare practitioners risk assessments for patients with intracranial haemorrhage and ultimately those patients' wellbeing. This analysis shows the severity of the impact of consent bias in the health field. Because a broad array of health measures are associated with consent, it is very likely that health linkage studies will be impacted by consent bias. If this bias goes unaccounted for, it can impact on study quality and ultimately research findings, medical best-practice, and policy.

Literature review: Consent related variables

There is a considerable amount of research that examines biases introduced by seeking consent. Most of this research investigates whether demographic, socioeconomic, and other variables are associated with an individual's consent decision. To date most of this research is in the health field. Understanding what participant characteristics are likely to impact on their consent decision is key to knowing whether consent bias is likely to impact on a study. For example, if a researcher knows that health status will impact on their participants' consent decision, they know that it is likely their research in health and ageing will be biased. This section reviews the literature on variables associated with consent to health data linkage, highlighting patterns and contradictions across studies.

Consent rates tend to be high, but differ across studies

Usually most people consent to having their data used for linkage research. Most studies have consent rates over 50%, many much higher. A systematic analysis of consent to data linkage found consent rates between 39% and 97% (da Silva et al., 2012). In their analysis of consent rates to epidemiologic studies, Dunn et al. (2004) looked at seven studies conducted by the Primary Care Sciences Research Centre at Keele University in the United Kingdom (UK). Across the studies a total of 42,812 people were asked for consent to link medical records, and overall 65% of them consented.

On a smaller scale, one study by Woolf, Rothemich, Johnson, and Marsland (2009) asked consent to link medical records as part of a survey. Of their 1106 participants, 67% consented, 25% actively declined, and 8% did not answer. Another survey of 5069 patients with asthma or angina in the UK obtained consent to access medical records from 61% of participants, with 9.8% of participants actively refusing consent (Baker, Shiels, Stevenson, Fraser, & Stone, 2000). One study of spinal cord injury using data linkage in Switzerland achieved an 80% consent rate across two modules. Among the groups that did not consent, 69% and 82% of those actively refused (Brinkhof, Fekete, Chamberlain, Post, & Gemperli, 2016).

In comparison to the large consent rates presented, The Avon Longitudinal Study of Parents and Children achieved a small 25% approximate overall participation. The authors sought consent to link the study data with participant's health, education, economic and criminal records. Of the 13,136 participants asked for consent, 28% responded. Of those, 81.2% consented to linkage, 17.2% consented to some linkages, and 1.6% declined (Audrey, Brown, Campbell, Boyd, & Macleod, 2016). The small consent rate in this study is largely accounted for by the small response rate.

Similarly Todd, Aitken, Boyd, and Porter (2016) received consent to linkage from 97% of their sample of mothers giving birth. This high consent rate is marred by a low participation rate; from a sample of 1989, only 913 responded to the survey. The authors point out that it is possible that the consent to linkage question reduced overall survey participation. This highlights a less direct way that seeking consent to link data can introduce bias. Unlike simply consenting or not consenting, declining to participate cannot be identified as a separate source of bias related to linkage.

Though limited in number, there are sufficient studies in New Zealand that offer an indication of consent rates for data linkage. In the Growing up in New Zealand study, researchers asked for consent to link children's longitudinal data while asking for consent to longitudinal research (Morton et al., 2010). Over 97% mothers consented to linkage, a substantial amount. This study is distinct from most consent-gaining scenarios in that it asks for consent to link someone else's data – the consenters' children.

This collection of studies demonstrates high but variable consent rates. This variability in consent highlights the importance of understanding not only consent rates but also consent-related variables across a variety of settings and studies.

Gender and age aren't clearly associated with consent

Age and gender are commonly incorporated as control variables and may mediate or moderate relationships investigated in health research. However, in studies specifically exploring consent to data linkage, neither variable has been consistently associated with consent to link data (Sakshaug & Kreuter, 2012). Some studies find that men consent at higher rates (Knies, Burton, & Sala, 2012; Woolf et al., 2009; Yawn, Yawn, Geier, Xia, & Jacobsen, 1998) while others find that women generally consent at higher rates (Dunn et al., 2004; Korkeila et al., 2001). Still other studies have found no association between gender and consent (Doyle & Sadler, 2013; Huang et al., 2007).

Regarding the influence of age on consent to data linkage, an overview of results from previous studies highlights similar variability as with participant age. Some studies find that older participants are more likely to consent to linking data. For example, the mean age of patients consenting to medical record linkage and further study was 41.3 years, whereas the mean age of those who did not consent was 38.8 years in one study (Woolf et al., 2009). In contrast, other studies have found either that older participants are less likely to consent, or that younger participants are more likely to consent, to linkage. A study in Taiwan found that those aged 65 and older were less likely to consent to data linkage (Huang et al., 2007) and increasing age was associated with reduced likelihood of consenting to medical record linkage in the UK (Dunn et al., 2004). Among participants

in the UK, those aged 16-24 were most likely to agree to have their data linked, with a 61% greater chance of consenting than participants aged 60 plus (Knies et al., 2012).

These conflicting results already have a tentative explanation. Age and gender may have a more complex inter-relationship with consent to medical record use and data linkage. When researchers account for both demographics, they have found that age and gender jointly influence consent. A 1996 study of consent to medical record usage found that women aged 41 to 64 years had an increased likelihood of actively declining to share their data compared with a high-consent reference group, who were younger and more likely to be men (Yawn et al., 1998).

Similarly, an analysis of seven surveys found that overall women consented to medical record linkage at higher rates than men until the 60-69 age bracket. Consent rates among women increased until age 40-49 and then declined, whereas consent rates among men increased until age 70-79 before decreasing (Dunn et al., 2004). This relationship remained in a logistic regression with other variables. This suggests that there may be a slightly different curvilinear relationship between age and consent to data linkage for each gender.

These differences by age and gender suggest that patterns of association with consent to link may be different in older adult samples. There has been no analysis of consent-related variables in an older adult sample to date.

Minority ethnicity can reduce consent rates

Minority ethnic status has consistently been found to be associated with lower consent rates. In the US, consenting participants are less likely to be African-American (Woolf et al., 2009). Similarly in the UK those living in England and with white heritage were more likely to consent to linkage than other groups (Knies et al., 2012; Mostafa, 2016).

When asking mothers in the UK to consent to linkage both on behalf of their children and themselves, Al Baghal (2016) found that Black, South Asian, and participants of other minority ethnicities had lower odds of consent for themselves and their children compared to white participants (OR 2.137 to 4.258, p<0.5). As a result, minority groups were under-represented in the data (Al Baghal, 2015).

This pattern seems somewhat consistent across many westernised countries, but it may not hold across all cultures. Many studies present conflicting ethnicity results (Kho et al., 2009). In Taiwan, Taiwan aborigines were 77% more likely to give consent than the majority participants after adjusting for other demographic variables (Huang et al., 2007). The authors suggested this may be

because of a lack of suspicion of healthcare research and considerable heterogeneity in the population.

There is one explanation for the rate of consent among minority participants that can account for findings by Huang, Shih, Chang, and Chou (2007): trust. The authors assert that Taiwan aborigines lack suspicion of healthcare research (Huang et al., 2007). Trust has been found to be positively correlated with consent when other variables are taken into account (Sala, Burton, & Knies, 2012). Al Baghal (2015) suggests that trust may influence consent decisions relating to ethnicity, socioeconomic status, age, length of study participation, rapport, and others. This highlights the importance of building trust with research participants. The ethnicity – consent relationship may be mediated by the level of trust in research by the ethnic group.

New Zealand is consistent with other westernised countries in that minority status does appear to be associated with reduced rates of consent. Carter, Shaw, Hayward, and Blakely (2010) examined variables associated with consent and found that participants of Māori, Pacific, and Asian decent had lower odds of consenting to linkage. The authors highlight concerns in New Zealand about participant burden among Māori and Pacific Island participants.

Māori and Pacific Peoples may have low trust in research (e.g. Bishop, 1998). Māori in particular have historically been both over-researched and in culturally inappropriate ways, leading to critique of western research and the creation of Kaupapa Māori research methods (Barnes, 2000; Powick, 2003; S. Walker, Eketone, & Gibbs, 2006). Māori may have concerns about how their data is used (J. Walker, Lovett, Kukutai, Jones, & Henry, 2017).

Socioeconomic factors have an unclear association with consent

The relationship between socioeconomic status (SES) and consent varies depending on both the study and measure used. Often studies will find one proxy measure of SES to be significant while others are not, but there is no pattern of any one SES measure being associated with consent consistently across studies. This lack of consistency in SES measurement presents a problem for determining whether, and how, SES is related to consent to link data.

Some studies have found indicators of higher SES to be associated with lower consent rates (e.g. Korkeila et al., 2001). Al Baghal (2015) found that children whose parents were on benefits were more likely to consent to linkage of their children's data as well as their own. Conversely, those parents with university degrees and who owned their own homes were less likely to consent to have linkage carried out with their children's data or their own data (Al Baghal, 2015).

Other studies have found indicators of increased SES to be associated with higher consent rates (e.g. Young, Dobson, & Byles, 2001) , or vis versa. Klassen, Lee, Barer, and Raina (2005) found higher family income to be associated with higher rates of consent to link data for parents with healthy babies. Huang et al., (2007) found that those who refused consent were more likely to be illiterate, have a lower monthly household income and live in suburban areas. Two studies have found consent to be associated with higher education but not other socioeconomic factors (Carter et al., 2010; Knies et al., 2012).

In contrast to previous research, which found either a positive or negative association between SES and consent, at least one study using occupation type as a proxy for SES found no association with consent (Doyle & Sadler, 2013). It appears that either there is no relationship between SES and consent, or that the relationship is not linear. Some research indicates a curvilinear relationship between consent and SES, with lower participation among those at the high and low ends of the socioeconomic distribution (Jutte, Roos, & Brownell, 2011).

Further research is needed to clarify this relationship between SES and consent. It is possible that the relationship only exists for some measures of SES, that other factors exacerbate or obscure it, or that it is non-linear. The current study includes several measures of SES that may shed light on the SES-consent relationship in this data.

Health is often significantly related to consent

Health status is often found to be associated with likelihood of consenting to data linkage. Most frequently, participants with poorer health or more health conditions are found to be more likely to consent. This appears to broadly hold true across different measures of health and research samples, with few exceptions.

Dunn et al. (2004) examined seven general population surveys in the UK conducted between 1996 and 2002, to which 25,000 people responded. They found a consistent association between health and consent across the studies. Participants reporting the symptom under investigation were 1.5 times more likely to consent to the researchers accessing their medical records than participants who did not report the symptom under investigation (Dunn et al., 2004).

Co-morbid health conditions have been found to have a strong association with higher rates of consent to linkage (Carter et al., 2010) as have lower overall health scores on the SF-36 (Woolf et al., 2009). Similarly there is a positive relationship with health conditions in children for whom consent is being sought (Klassen et al., 2005). Specific health conditions found to be positively associated with consent include panic disorder (Korkeila et al., 2001). Health behaviours found to have a

positive association with consent include smoking, using tranquilisers, and drinking more alcohol (Korkeila et al., 2001).

In contradiction to Woolf, Rothemich, Johnson, and Marsland (2009), other authors illustrated the importance of the statistical technique by reporting both bivariate and multivariate associations with consent. They found that once other variables had been accounted for, none of the SF-36 health status measures had an impact on likelihood to consent to data use for research purposes (Huang et al., 2007).

Poorer health does not always result in higher rates of consent, occasionally the relationship is reversed. Young, Dobson, and Byles (2001) found that women consenting to data linkage were less likely to experience major personal illness, hospitalisation in the previous year, and had lower use of GP services than non-consenters. In addition, "the mortality rate among older women was higher for non-consenters than consenters" (p. 418). The authors make no comment on how their findings contradict the literature on consent to health linkage generally.

Observational research on heart disease revealed four factors as being associated with consent: having had artery widening surgery, lower blood pressure, lower cholesterol level, and being an ex-smoker (Buckley et al., 2007). The authors suggested that the consent trends they found were indicative of participants who pursue healthy lifestyles and benefit from healthcare intervention consenting at higher rates than others.

It is still unclear why health status might impact on a participant's decision to consent to health linkage. However, enough theory focussed research has been done to make speculation possible. Participants in poor health may find the request more salient than healthy participants, particularly where they suffer from the symptom being investigated (Dunn et al., 2004; Mostafa, 2016; Sala et al., 2012). Alternatively, those with serious health conditions may have a different perspective on the privacy risks associated with participation. There is a clear association between disposition towards being private and reduced consent (Mostafa, 2016; Sala et al., 2012).

Finally, participants in poorer health may feel that they have more to contribute or are socially obligated to participate because of the social good associated with linking health data (Millican & Mansfield, 2013). This is suggested by Sala, Burton, and Knies' (2012) finding that community mindedness is strongly associated with consent. Whatever the reason for this bias, over-representation of participants with poorer health is likely to influence findings in many areas of health research. As such it is a key concern for researchers using linked health data.

More and longer participation increases consent rates

Most studies of consent to linkage only examine demographic, health, and socioeconomic characteristics of the individual consenting. However, newer studies are emerging that also cover other domains, including the individual's relationship to the study. Some studies suggest that greater participation in the survey increases subsequent likelihood of consenting to having that survey data linked. Conversely, decreased participation or resistance to participation lowers the likelihood that a person will consent.

Only a few studies directly examine length of participation or number of instances of participation, but to date they have found similar results. One study found that participants who had missed one or more of the four survey waves were significantly less likely to give consent to link their own or their child's data (Mostafa, 2016). Another survey, examining different methods of consent-seeking found that participants asked to consent in later waves were more likely to give consent than participants asked in earlier waves (Sala, Knies, & Burton, 2013).

Despite previous research showing a positive association between length of survey participation and consent, it is unclear why the association exists. Potential explanations include trust in the research (Sala et al., 2013) or a general tendency to participate (Sakshaug et al., 2016). The latter appears to be supported by several findings. In one postal survey, participants who did not need survey reminder letters were more likely to consent to linkage (Klassen et al., 2005). Similarly the more times participants in a telephone-based survey were called before participating, the less likely they were to consent to having their data linked (Al Baghal, 2015).

Along similar lines Sakshaug, Couper, Ofstedal, and Weir (2016) had interviewers assess participant resistance to participation and correlated this with consent to data linkage. They found that participants assessed as more uncooperative were less likely to consent. Alone this finding is unreliable. The interviewer is likely to be a confounding factor, both assessing the resistance of the individual and asking their consent. Interviewer characteristics have been found to be associated with consent to linkage (Korbmacher & Schroeder, 2013).

However, the authors went a step further and found that participants who initially declined to participate were also less likely to consent to linkage (Sakshaug et al., 2016). Together these findings suggest a tendency for individuals to want to participate or not, which may increase or decrease the likelihood of consent to linkage even if the participant has agreed to participate in the main survey.

Consent bias for older adults hasn't been researched

It has clearly been established that some factors (e.g. health status, ethnicity, etc.) may influence the degree to which participants consent to data linkage. However, most of this research has been conducted on the general population. It is unclear to what extent these findings may be relevant to a sub-population increasingly asked to consent to long-term health studies which increasingly include data linkage: older adults.

The last 20 years has witnessed a rapid increase in population ageing (United Nations Department of Economic and Social Affairs Population Division, 2015) and a parallel increase in studies of ageing that seek to identify the factors critical to maintaining health and independence in later life. A few examples include the US Health and Retirement Study (Heeringa & Connor, 1995), the English Longitudinal Study of Ageing (Taylor et al., 2007), Survey of Health Ageing and Retirement in Europe (Bo¨rsch-Supan, Hank, & Ju¨rges, 2005), the Korean Longitudinal Study of Ageing (Park et al., 2007), and the New Zealand HWR (Dulin et al., 2011; Towers et al., 2017).

Many of these studies seek to undertake ongoing population surveys of health, wealth, and social interaction and to link the resulting data with existing health administrative datasets to understand the influence of these factors on healthcare utilisation. Given the rapid rise in the use of data linkage in studies of health and ageing, and the impact that consent bias can have on study outcomes, it is essential that we understand the possible factors influencing consent to data linkage. Understanding what variables are related to consent for older adults will enable researchers to minimise, account for, and avoid the effects of consent bias in health and aging research.

Some researchers examining variables influencing consent have mentioned consent among older adults, but only as part of samples including adults of all ages. Older adults may generally consent at a higher rate (Carter et al., 2010; Huang, Shih, Chang, & Chou, 2007; Woolf, Rothemich, Johnson and Marshland, 2000). Other research has found gendered differences in consent at different ages. Dunn et al. (2004) investigated several UK samples. They found that response and consent rates increased until age 70 for women and age 80 for men, at which point consent rates dropped (Dunn et al, 2004). While this does highlight some consent differences within an older adult sample, no research to date has focused on older adults specifically to identify systematic differences between consenters and non-consenters within this group.

Research Aims

This study seeks to fill a gap in the literature by examining variables related to consent amongst older adults. To do this, it will make use of the Health, Work and Retirement Longitudinal Study

(HWR; Towers et al., 2017). This is a longitudinal survey that began in 2006, examining health and ageing amongst older New Zealanders. The HWR study examines a variety of variables including physical and mental health, various demographic characteristics, quality of life, social support and networks, work status, work stress, care giving, and safety issues. The original HWR sample remaining in 2014 consisted of 1,345 community dwelling, ageing New Zealanders.

In 2014 participant consent was sought to link the HWR longitudinal data with data sets held by the Ministry of Health. Participants were sent a letter inviting them to join this data linkage along with a consent form and information pack. Of the participants who were sent consent forms, roughly half consented to the data linkage. The remainder either did not respond or responded and declined to provide consent.

The consent process provided a unique opportunity to examine what demographic, health, and personality variables may have an influence on older adults' choice to consent to data linkage. Research examining variables related to consent, and consent bias, is difficult. Researchers who begin with a data linkage project often do not gather data prior to asking consent to linkage, or consent requirements do not allow for the use of such information. The data from non-consenters is often lost. Without non-consenters' data it becomes impossible to examine differences between consenters and non-consenters and generalise to the wider population (Huang, Shih, Chang & Chou, 2007). In the case of the HWR linkage, even those participants who have not consented to the data linkage have given consent for their existing data to be used for research within the scope of the study aims, enabling the use of data from both linkage consenters and non-consenters for a project of this kind.

This thesis seeks to answer the question, 'To what degree are various demographic and health variables related to consent to data linkage among older adults in New Zealand?' Demographic variables previously found to be related to consent to linkage will be examined in the original HWR sample; including age, sex, education, income, socioeconomic status, and health status. Based on the findings of research explicitly covered in this introduction, the hypotheses for the current research are:

H1. Participants of Māori and Pacific Islander ethnicity will have lower rates of consent than New Zealand Europeans.

H2. Broad measures of health such as SF-12 and number of health conditions will have a negative relationship with consent, where poorer health will be associated with higher rates of consent.

H3. Being involved for a greater number of waves will be associated with higher consent rates.

Previous research does not provide sufficient evidence for hypotheses about age, gender, or specific health issues. Research findings about age and gender are conflicting, while findings about specific health conditions are rare. Because the HWR data includes measures of individual health issues, the current study will investigate their relationship to consent in this sample to add to the literature relating specific health issues to consent.

Method

Participants

This study is a secondary-analysis of data from an existing longitudinal study. Participants for this study were drawn from those who had completed the 2013 data collection wave of the Health Work and Retirement Longitudinal Study (HWR; Towers, Stephens, Breheny, & Allen, 2016). Participants in the HWR study were originally drawn from the New Zealand electoral role in 2006 using equal probability random sampling. A Māori over-sample was created using the self-report Māori descent indicator on the general electoral roll to enable more reliable analysis of differences between Māori and non-Māori wellbeing. Of the initial sample in 2006, 3,127 adults aged 55-70 agreed to participate in the longitudinal study (Towers, Stephens, Breheny, & Allen, 2016). Attrition over time has resulted in the 2013 HWR sample including some total of 1,345 community dwelling participants.

Those HWR participants recruited in 2006 who had completed the 2013 HWR data collection wave were subsequently approached in 2014 for consent to link their longitudinal research data with their National Health Records. These records are held on National Medical Information Databases by the New Zealand Health Information Service (NZHIS), part of the Ministry of Health. The five NZHIS databases linked with the HART longitudinal data were:

- **National Minimum Dataset**: **Hospital Events**, which contains hospital discharge information including clinical information (e.g. health diagnosis) and event information (e.g. times and dates)

- **National Non-Admitted Patient Collection,** which contains data about non-admitted face-to-face secondary care events

- **New Zealand Cancer Registry**, which contains all primary cancers diagnosed in New Zealand

- **Pharmaceutical Collection**, which contains information from pharmacists that is used for subsidised dispensing

- **Mental Health Information Collection**, which contains mental health information including care provided, diagnosis of mental health condition, legal status, and discharge

Participants who had completed the 2013 HWR data collection wave were sent information sheets and consent forms which outlined the nature of the request, the specific datasets for which access was sought, and their rights to consent or not to consent. Participants who consented to this data-linkage were asked to sign and return the consent form in a freepost envelope provided. Ethical approval for this data linkage was provided by both the Massey University Human Ethics Committee

(HEC: Southern A – 13/62) and the New Zealand Health and Disability Ethics Committee (14/CEN/79). Of the HWR participants who participated in the 2013 data wave, 1.3% (18) were not approached for consent. These participants had withdrawn (3), were deceased (14), or had lost contact (1). These participants were removed from the analysis.

Of the 1,327 participants approached for consent, 781 (58.9%) consented, 543 (40.9%) did not respond and 3 (0.2%) actively declined. Because the group of participants actively declining was so small, non-responders and decliners were analysed as one group resulting in two categories: consenters and non-consenters. This facilitated the use of the 2013 HWR dataset as a foundation on which to assess the key health, wealth, social and demographic characteristics differentiating those who would subsequently consent to data linkage from those who would not.

Table 1 below provides an overview of the final sample for this study. In 2013 these participants were aged 60-77 with a mean age of 68.52 (*SD* = 4.471). Just over half the participants—721 (54%)—were female. The study included a Māori oversample, resulting in 41% participants with Māori ancestry; although when asked what ethnicity they considered themselves, only 33.3% prioritised Māori. The sample was evenly distributed in their qualifications: 21.9% had no qualifications, 19.7% had completed a secondary school qualification, 27.1% had obtained a post-secondary or trade qualification, and 29.9% held a tertiary qualification. A total of 73 participants chose not to provide income information.

Table 1
An Overview of the 2013 HWR Sample

	N	(%)
Total N	1327	100%
Gender		
Male	606	45.7%
Female	721	54.3%
Prioritised Ethnicity		
NZ European	784	59.1%
Māori	442	33.3%
Pacific peoples	78	5.9%
Asian	4	0.3%
Other (including non-response)	19	1.4%
Highest Educational Qualification		
No qualifications	291	21.9%
Secondary school	261	19.7%

Post-secondary/trade	359	27.1%
Tertiary	397	29.9%
Net personal income*		
< $20,000	543	40.9%
$20,000 - $39,000	377	28.4%
$40,000 - $59,000	196	14.7%
$60,000+	138	10.4%
ELSI-SF Category Score		
Severe hardship	36	2.7%
Significant hardship	37	2.8%
Some hardship	66	5.0%
Fairly comfortable	122	9.2%
Comfortable	240	18.1%
Good	471	35.5%
Very good	287	21.6%

Note: N for each column may not reflect total N (1,327) due to missing data on some variables.

* The 11 original income categories were collapsed into 4 for ease of reporting in this table.

Measures

Consent. Consent was measured by participants' response to the request for consent to data linkage sent after the HWR 2013 data collection wave. After data collection in 2013 all respondents to that dataset were sent a letter inviting them to consent to have the HWR team access their National Health Records. This allowed the team to link that data with the survey data already on file from previous HWR data collection waves. Responses to this request fell into three categories: Consent, Decline, or No Response. Due to the number of participants in the Decline group (N=3), the Decline and No Response categories were collapsed into a single category. This resulted in a dichotomous consent indicator; Consented and Did Not Consent.

Demographics. Participant age was indicated in years. Participants were asked to indicate all ethnicities to which they belonged including New Zealand European, Māori, Pacific Peoples, Asian, MELAA, or Other, specified in writing by the participants. Ethnicity responses were then 'prioritised' following a procedure previously established by Statistics New Zealand (Statistics New Zealand, 2004) to identify a prioritised ethnic group from those outlined. Because so few participants indicated Asian, MELAA or Other as their prioritised ethnicity, these categories were collapsed into

an overall Other category for analysis purposes. Income was self-reported by participants in New Zealand dollars and then collapsed into 11 categories with $10,000 intervals.

Socioeconomic status. Socioeconomic status was measured using two indicators that capture area level deprivation and individual living standard respectively, which are different but complementary aspects of socioeconomic status. The NZDep2006 (Salmond, Crampton, & Atkinson, 2007a) measures the general socioeconomic deprivation of an area by combining nine socioeconomic variables from the 2006 Census. NZDep2006 scores are categorised into deciles from 1 (least deprived) to 10 (most deprived). To determine the NZDep2006 score, participant addresses are geocoded. Deprivation scores are then determined based on the NZDep2006 dataset (Salmond, Crampton, & Atkinson, 2007b). While the NZDep2006 does not measure individual-level deprivation, it is a useful estimate of deprivation. It also accounts for neighborhood effects of deprivation when considering how deprived an individual is.

The ELSI$_{SF}$ (Jensen, Spittal, & Krishnan, 2005) is a shortened version of the ELSI, which captures a full range of living standards by measuring consumption items, amenities, and recreations which typically are available or lacking at different standards of living (Jensen, Spittal, Crichton, Sathiyandra, & Krishnan, 2002). The scale is made up of four aspects – ownership and social participation restrictions, economising, and self-rating scales (Jensen et al., 2005). ELSI$_{SF}$ scores range from 0 to 31 and are divided into seven levels with the labels Severe Hardship (indicating respondents lack 39% of basics, have 10% of luxuries, and have 52% of the financial problems), Significant Hardship, Some Hardship, Fairly Comfortable, Comfortable, Good, and, Very Good (indicating respondents lack 0% of basics, have 88% of luxuries and have 1% of the financial problems). Alternatively, ELSI scores can be divided into three categories; Hardship, Comfortable, and Good. ELSI$_{SF}$ scores were measured in the 2013 wave.

Health status. Health status in this study was represented by two specific variables: the SF-12 and a checklist of chronic health conditions. The SF-12 (Ware, Kosinski, & Keller, 1996) is a shortened version of the SF-36 health survey. Respondents recall the preceding four weeks to answer twelve questions that cover eight areas of health: general and mental health, vitality, social functioning, role limitation due to physical health and emotional problems, bodily pain limiting activities, and physical functioning. The Physical Component Summary scale (PCS) and Mental Component Summary scale (MCS) summarise these dimensions.

Ware, Kosinski, and Keller (1996) found a test–retest correlation of 0.89 for the PCS and 0.76 for the MCS. Summary scores range from 0 (poorest) to 100 (best) score, with higher scores indicating

better health on each subscale (Cernin, Cresci, Jankowski, & Lichtenberg, 2010). The internal consistency of the SF-12 among an older adult sample was a=0.84 (Resnick & Nahm, 2001).

The 2013 HWR data collection wave was principally focused on exploring psychosocial factors in ageing and, as such, has limited health data collection other than the SF-12. However, the 2012 HWR data collection which had occurred approximately 12 months prior to the 2013 data collection wave had included a checklist of chronic (i.e. significant and on-going) health conditions. This chronic health condition data was integrated with the 2013 data and used as an indicator of on-going chronic health problems for the purposes of this study.

Participants in 2012 were asked whether they had any of the following health problems: arthritis or rheumatism, disorder of the neck or back, diabetes, disability, hearing impairment, heart trouble, high blood pressure or hypertension, mental illness, respiratory condition, sight impairment, sleep disorder, stroke, cancer and cancer type, or another condition. Participants indicated whether they had each condition using a dichotomous 'Yes' or 'No' indicator. A proxy indicator of total health conditions experienced was created by summing all 'Yes' responses across each condition. The creation of this variable allowed comparison with other linkage studies that use number of health conditions as a health measure.

Analysis

Analyses in this study were conducted using SPSS version 24. Several variables were re-coded to enable the analysis and ease of interpretation. The health problem variables were recoded so that 0 = no and 1 = yes. The ELSI category scores were re-coded so that lower numbers represented poorer socioeconomic status and larger numbers represented improved socioeconomic status.

Bivariate analysis. To determine the bivariate relationship between individual variables and consent, chi-squared tests were completed with each categorical variable and t-tests were conducted with each continuous variable. Before completing bivariate analysis, some variables were re-coded to ensure that the chi-square tests met the assumption of at least five observations in each cell, and that the categories were easy to interpret. Income categories were recoded from 10 categories of $10,000 increments up to $100,000+ into six categories of $20,000 increments, up to $100,000+. Number of health conditions was recoded into four categories relating to the number of reported health conditions, from 0 to 3+.

Multivariate analysis. A binary logistic regression was chosen as the most appropriate statistical test to use for the multivariate analysis. Binary logistic regression enables the analysis of the impact of multiple independent variables on a single, dichotomous dependent variable. In this case the

dependent variable is consent, with two levels; consented and did not consent. Multiple continuous and categorical independent variables were used in the current study.

Categorical independent variables must be dummy coded to enable them to be included in the analysis. Because of the way the variables were coded, the dummy coding always used 0 as the reference category which always referred to the absence of the measured variable or poorest outcome. Assumption testing for the model is described in the results section.

Results

Table 2 provides a demographic breakdown of the sample in total and differentiated by whether individuals chose to consent for data linkage. Additionally, Chi-squared results and effect sizes are reported.

Table 2
Demographic information for consenters and non-consenters

	N (%)	Consented	Did not consent	χ^2	ϕ
Total N	1,327 (100.00%)	781 (58.85%)	546 (41.15%)		
Gender					
Female	721 (54.33%)	401 (51.34%)	320 (58.61%)	6.54*	0.07
Male	606 (45.66%)	380 (48.66%)	226 (41.39%)	-	-
Prioritised Ethnicity					
NZ European	784 (59.76%)	506 (65.29%)	278 (51.77%)	25.84**	0.14
Māori	442 (33.69%)	220 (28.39%)	222 (41.34%)	-	-
Pacific peoples	78 (5.95%)	45 (5.81%)	33 (6.15%)	-	-
Other (including non-response)	23 (0.61%)	10 (0.52%)	13 (0.74%)	-	-
Highest Educational Qualification					
No qualifications	291 (22.25%)	154 (19.87%)	137 (25.70%)	12.18**	0.10
Secondary school	261 (19.95%)	144 (18.58%)	117 (21.95%)	-	-
Post-secondary/trade	359 (27.45%)	233 (30.06%)	126 (23.64%)	-	-
Tertiary	397 (30.35%)	244 (31.48%)	153 (28.71%)	-	-
Net personal income					
< $19,999	543 (43.30%)	319 (42.14%)	224 (45.07%)	3.95	0.06
$20,000 - $39,000	377 (30.06%)	233 (30.78%)	144 (28.97%)	-	-
$40,000 - $59,000	196 (15.63%)	120 (15.85%)	76 (15.29%)	-	-
$60,000 - $79,000	76 (6.06%)	47 (6.21%)	29 (5.84%)	-	-
$80,000 - $99,000	18 (1.44%)	14 (1.85%)	4 (0.80%)	-	-
$100,000 +	44 (3.51%)	24 (3.17%)	20 (4.02%)	-	-
ELSI-SF Category Score					
Severe hardship	36 (2.86%)	17 (2.29%)	19 (3.67%)	21.71**	0.13
Significant hardship	37 (2.94%)	16 (2.16%)	21 (4.05%)	-	-
Some hardship	66 (5.24%)	26 (3.51%)	40 (7.72%)	-	-
Fairly comfortable	122 (9.69%)	71 (9.58%)	51 (9.85%)	-	-
Comfortable	240 (19.06%)	146 (19.70%)	94 (18.15%)	-	-

Good	471 (37.41%)	277 (37.38%)	194 (37.45%)	-	-
Very good	287 (22.80%)	188 (25.37%)	99 (19.11%)	-	-
Total number of health conditions				
0	229 (17.26%)	125 (16.01%)	104 (19.05%)	3.05	0.05
1	259 (19.52%)	155 (19.85%)	104 (19.05%)	-	-
2	279 (21.02%)	160 (20.49%)	119 (21.79%)	-	-
3+	560 (42.20%)	341 (43.66%)	219 (40.11%)	-	-

Note: Not all participants responded to some demographic variable questions (e.g. net personal income) and so these may not add up to 1,327. Significant differences between the two consent-based groups as determined by Chi-square (χ2) test. Effect size for between-group difference indicated by ϕ (Phi coefficient), values around 0.1 indicate a small effect size, values around 0.3 indicate a medium effect size, and values around 0.5 indicate a large effect size.

*$p = <0.05$, ** $p = 0.01$

Of the categorical variables analysed only gender, ethnicity, highest educational qualification and ELSI-SF category score had a significant relationship with consent. Effect sizes varied from 0.07 (gender) to 0.13 (ELSI-SF category score), indicating a small effect size. While these variables are related to consent, and that relationship is unlikely to be by chance, they only have a small effect on consent.

In addition to the categorical indicators included in Table 2, independent samples t-tests were conducted to compare consenters and non-consenters across continuous variables included in this study. These are presented in Table 3 below.

Table 3

Comparison of consenters and non-consenters across continuous variables

	Consenters		Non-consenters					
	M	SD	M	SD	t-test	Mean difference	95% CI	Cohen's d
Age	68.73	4.58	68.23	4.29	-2.04*	-0.50 *	-0.99 – -0.01	-0.11
NZ Dep score	977.59	81.04	996.27	97.38	3.68**	18.68 **	8.72 – 28.65	0.21
SF-12 PH	48.50	10.42	47.70	10.14	-1.36	-0.81	-1.97 – 0.36	-0.08
SF-12 MH	49.56	7.84	49.31	8.00	-0.55	-0.25	-1.14 – 0.65	-0.03
CES-D-10	5.76	4.82	6.29	4.66	1.95	0.53	-0.00 – 1.06	0.11
No. of waves participated in	5.80	0.44	5.58	0.68	-6.67**	-0.22**	-0.28 – -0.16	-0.40

Note: Cohen's d values around 0.2 indicate a small effect size, values around 0.5 indicate a medium effect size, and values around 0.8 indicate a large effect size. M = Mean, SD = Standard Deviation, CI = Confidence Interval.

*$p = <0.05$, ** $p = <0.01$

Of the six continuous variables analysed, only three of them were significantly related to consent. They were age, NZ Dep score, and number of waves participated in. Age and NZDep score each had

small effect sizes. However, number of waves participated in had a medium effect size. This suggests that of all variables found to be related to consent so far, the number of waves a person had participated in was associated with the biggest difference in consent. Regardless, this effect is still not large.

Health conditions and likelihood of consent

Individual health condition variables were analysed to determine whether they were related to consent. Figure 1 illustrates the proportion of those with or without each condition that chose to consent to data-linkage. Most health conditions were not significantly related to consent.

However, two health conditions were significantly related to the likelihood of consent: diabetes and cancer. Notably the two conditions have contrasting relationship; cancer appears to increase the likelihood of consent while diabetes decreases it. Only 47.3% of participants with diabetes (70/148) consented to data linkage, compared with 60.3% of those without diabetes (711/1179), a difference that was statistically significant $\chi2$ (1, n = 1327) = 8.659, p = 0.003.
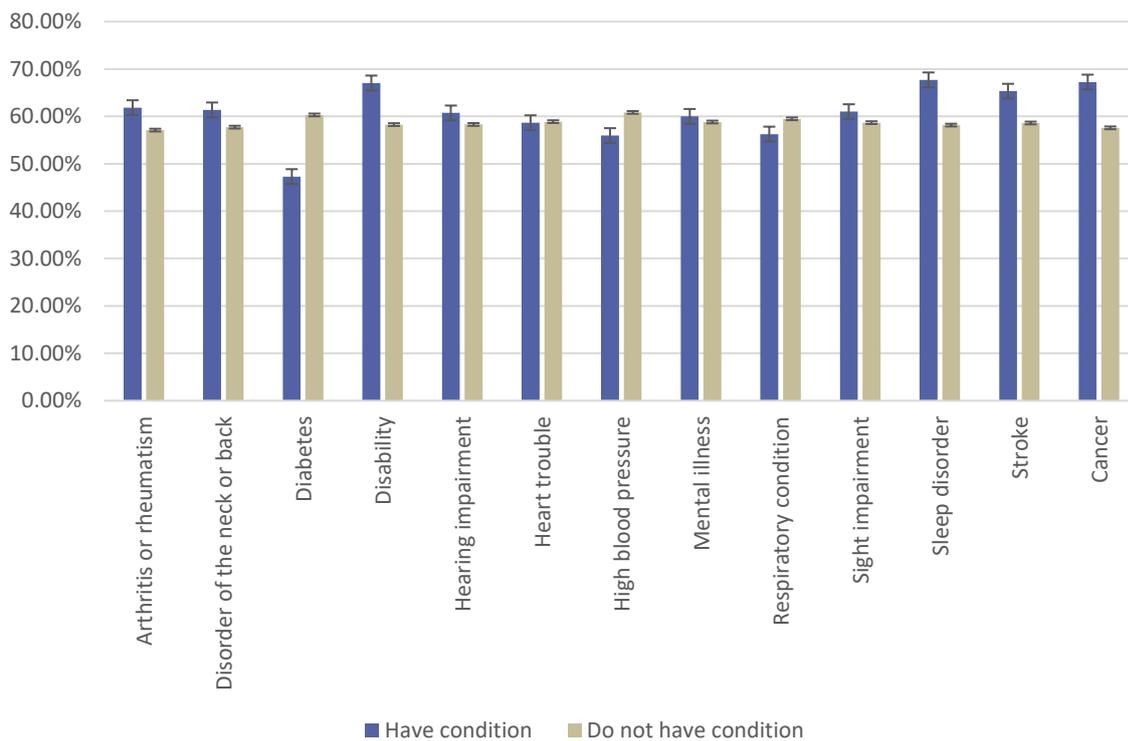

Figure 1. Proportion of Data-Linkage Consenters by Chronic Condition

Of participants with cancer, 67.2% (117/174) consented to have their data linked while only 57.6% of those without cancer consented to linkage (664/1153), $\chi2$ (1, n = 1327) = 5.425, p = 0.020.

Binary logistic regression

To determine what variables are related to consent in a multivariate analysis, a binary logistic regression was conducted. Logistic regression analysis was performed using SPSS statistical software.

From the bivariate analyses, the significant predictors of consent were gender, prioritised ethnicity, highest educational qualification, ELSI-SF category score, diabetes and cancer, age, NZ Dep Score, and the number of waves a participant had been involved in. From these variables, the logistic regression model was created.

ELSI-SF category score and NZ Dep Score were both retained in the regression model. Although they are both measures of socioeconomic status, they measure different aspects of it. The ELSI-SF is a person-level measure of the individual's means, capturing the full socioeconomic continuum (Jensen et al., 2002). Conversely, the NZ Dep is an area-level measure, applying to the context in which the individual lives. It is also focused on deprivation, capturing only the degree to which people in an area lack socioeconomic resources (Salmond et al., 2007a).

Some association should be expected between the two measures, so a bivariate correlation was performed with the ELSI-SF score and NZ Dep score to ensure that any association wasn't great enough to introduce multicollinearity to the model. The association was statistically significant but small $r(1256) = -0.25$, $p < 0.000$, as expected. It should be noted that the model uses the ELSI-SF categories rather than raw score.

The resulting model included nine variables: age, gender, prioritised ethnicity, highest educational qualification, diabetes, cancer, NZ Dep score, ELSI-SF category score, and number of waves participated in. Figure 2 depicts the model visually.
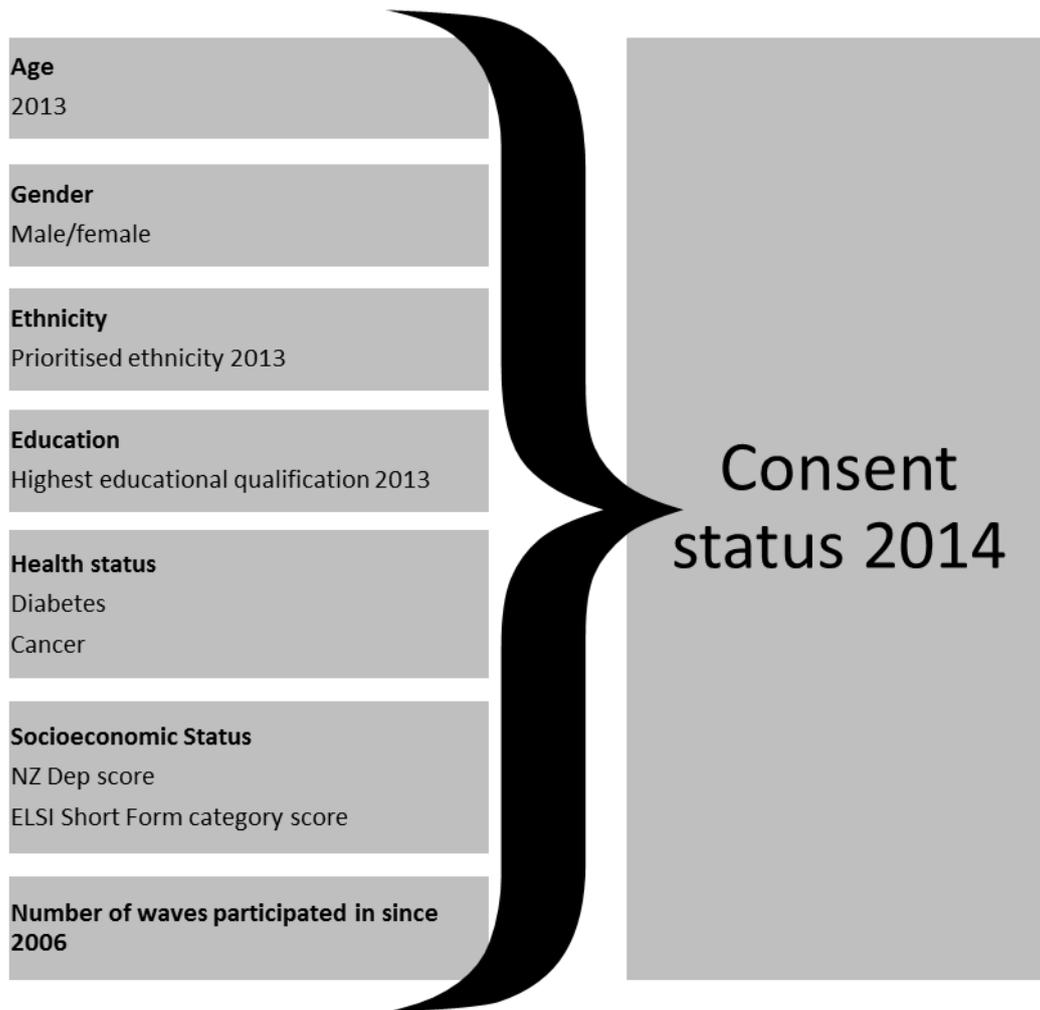
Age 2013	
Gender Male/female	
Ethnicity Prioritised ethnicity 2013	
Education Highest educational qualification 2013	Consent status 2014
Health status Diabetes Cancer	
Socioeconomic Status NZ Dep score ELSI Short Form category score	
Number of waves participated in since 2006	

Figure 2: A visual representation of the logistic regression model

Assumption testing. Prior to undertaking the regression analysis, the data was checked to ensure that it met the assumptions for a binary logistic regression. Multicollinearity was assessed by checking the correlations between continuous variables. No strong correlations were found. The standardised residuals were examined to determine whether outliers were present. The highest value was -2.16, suggesting that there were no outliers. The residuals plot in Figure 3 plots the predicted vs. observed probabilities in the model (residuals) against the predicted probabilities. This graph firstly reinforces the fact that there are no major outliers impacting on the model (Coolican, 2009). Additionally, the relative symmetry of the points indicates that linearity can be assumed.
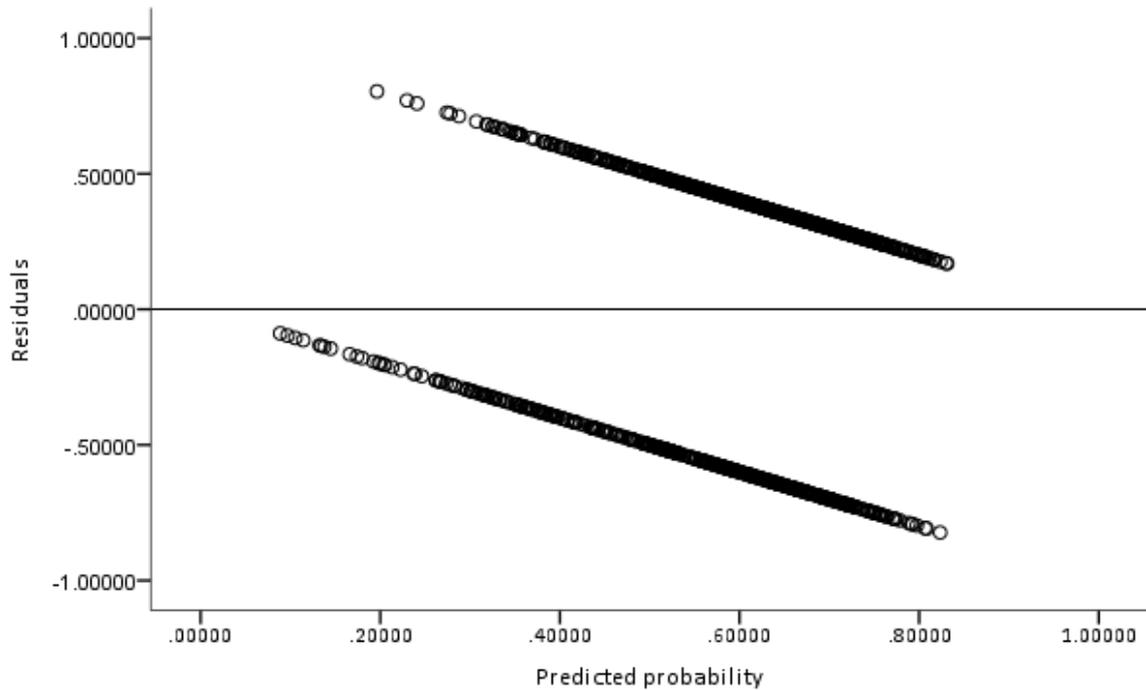
*Figure 3:*Residual plot for the binary logistic regression model

Results of the model. Of the 1,327 individuals whose data was included, 86 (6.48%) were excluded from the analysis because of missing data, leaving 1,241 cases in the logistic regression. The Hosmer Lemeshow goodness of fit test was not significant (p = 0.74) indicating that the model was well fitted. The regression model was statistically significant χ2 (18, n = 1241) = 94.90, p < 0.00, and correctly predicted 63.26% of cases, with a sensitivity of 85.42% and a specificity of 31.16%. These statistics indicate that, while the model was good at successfully predicting consenters, it was poor at predicting non-consenters.

Odds ratios for each variable in the regression are included in Table 4 below. The odds ratio (OR) requires careful interpretation (Osborne, 2006) and is interpreted slightly differently for categorical variables than it is for continuous variables. An OR greater than 1 indicates a positive association (i.e. a higher likelihood of occurrence than the reference group), with an odds ratio less than indicating 1 a negative association (i.e. a lower likelihood of occurrence). Of the variables included in the model, three significantly contributed to predicting consent – Māori prioritised ethnicity, having diabetes, and being involved for more waves.

Table 4

Odds ratios for variables in the binary logistic regression

	OR	95% CI for OR	
Categorical variables		Lower	Upper
Gender			

Female	1.00	-	-
Male	1.22	0.95	1.55
Ethnicity			
NZ European	1.00	-	-
Māori	0.68**	0.52	0.88
Pacific Peoples	0.87	0.52	1.44
Other	0.60	0.11	3.21
Highest qualification			
No qualifications	1.00	-	-
Secondary school	0.88	0.61	1.27
Post-secondary / trade	1.37	0.97	1.94
Tertiary	0.98	0.70	1.38
Health condition			
Diabetes	0.66*	0.45	0.96
Cancer	1.38	0.96	1.98
ELSI-SF category			
Severe hardship	1.00	-	-
Significant hardship	1.02	0.38	2.72
Some hardship	0.77	0.33	1.82
Fairly comfortable	1.52	0.70	3.31
Comfortable	1.57	0.76	3.28
Good	1.40	0.69	2.84
Very good	1.82	0.87	3.80
Continuous variables			
Age	1.02	0.99	1.05
NZ Dep score	1.00	1.00	1.00
Number of waves	1.88**	1.50	2.36

*p = <0.05, ** p = <0.01

For categorical variables, the OR indicates the likelihood of being in the consent group relative to the reference category. The results in Table 4 indicate that the odds that someone with diabetes will consent to data linkage are 0.66 times those without diabetes, and the odds that Māori participants will consent are 0.68 times those that are New Zealand European. This means that older participants without diabetes are more likely to consent to data linkage than those with diabetes, while older New Zealand European participants are more likely than older Māori participants to consent.

For continuous variables, the interpretation of the odds ratio is slightly different. The odds ratio refers to the change in unit rather than comparing categories. The results in Table 4 regarding continuous variables indicate that for each additional wave of involvement a person's odds of consenting is 1.88 times the odds of consenting in the previous wave of involvement. This means that being involved for more waves increased older participants' likelihood of consent to data linkage.

While these factors were significantly related to consent in the multivariate model, the model only explained a small proportion of the difference between consent and non-consent. The Cox & Snell R Squared and Nagelkerke R Square, of 0.07 and 0.10 respectively, indicate that the model explained between 7% and 10% of a participant's decision to consent to linkage.

Discussion

This study examined consent bias for older adults

Worldwide large, longitudinal studies of ageing have started to link survey data to individual health records. Several studies have explored the key factors characterising willingness to consent to data linkage and how this might lead to biases in results in the general population. However, it is unclear if older study participants share the same characteristics as younger people and whether such characteristic differences between older consenter and non-consenters might result in result biases. This study used the 2013 dataset from the Health, Work and Retirement (HWR; Towers et al., 2017) longitudinal study to assess whether those who consented to a request to link their data significantly differed to those that did not consent and, if so, how such difference might potentially bias resulting analyses.

Māori ethnicity was negatively related to consent

The first hypothesis in this study was that participants of Māori and Pacific Islander ethnicities would have lower rates of consent than New Zealand Europeans. The study results partially supported this hypothesis. In the bivariate analysis, consent rates were lower for both older Māori and Pacifica older adults. In the multivariate model, older Māori had 0.68 times the odds of consenting and older Pacific Peoples 0.87 times the odds of consenting compared with older New Zealand Europeans. This was only significant for Māori participants. Māori ethnicity was one of only three variables in the multivariate analysis that significantly contributed to distinguishing between consent and not consent.

This finding is generally consistent with previous research showing an association between indigenous and minority ethnicity and consent, for both data linkage and research in general. Studies in the US (Woolf et al., 2009), the UK (Al Baghal, 2015; Knies et al., 2012), and previously in New Zealand (Carter et al., 2010) have all found that minority ethnicity status is associated with a lowered consent rate. These prior studies used samples that were either younger than the current sample or representative of their wider populations, including people of all ages. The current study expands this finding to show that indigenous ethnicity also influences consent in samples consisting only of older adults.

It is unclear why Pacific Peoples and other minority New Zealand ethnic groups did not also have a significantly decreased likelihood of consent. It is possible that the current study did not have enough participants from these ethnic groups to identify a significant association. Only 78 Pacific Peoples were involved in the study, and only 23 individuals with an 'Other' ethnicity were included.

Future research in New Zealand should further explore whether other minority groups are less likely to consent to data linkage.

Systematic differences in consent to health research linkage by ethnicity are concerning for several reasons. It may result in under-representation of health conditions that are more prevalent in minority populations (Al Baghal, 2015), leading to underestimation of the problem posed by those health conditions. It may also prevent researchers from being able to fully understand the health challenges of minority groups.

This is relevant in New Zealand as it is in other countries where indigenous and minority populations suffer poor health outcomes. Despite a narrowing of the life expectancy gap, Māori still die on average 7.1 years earlier than non-Māori (Macpherson, 2015). Older Māori have poorer health outcomes and a higher burden of chronic illness than older non-Māori (Ministry of Health, 2011). Māori and Pacific Peoples experience a number of other poor health outcomes including higher rates of diabetes, obesity, disability, cardiovascular disease, and a greater unmet need for mental health services (Ministry of Health, 2015, 2016).

A lack of trust in research is a plausible explanation for the decreased consent rate among some ethnic groups. Internationally there is a history of unethical, exploitative, and harmful research practices with minority populations, likely leading to mistrust among those ethnic groups. The Tuskeege Syphilis experiment is a particularly infamous and extreme example of the harms inflicted on disempowered minority ethnic groups though research (Lombardo, 2006; Paul & Brookes, 2015). More commonly, indigenous populations have been denied access to or control over data about their own health (J. Walker et al., 2017).

In New Zealand researchers similarly have often not taken the needs, tikanga (practices), or data ownership of Māori into account when undertaking research with them (Bishop, 1998; J. Walker et al., 2017) and Māori knowledge has not been valued the same way as Pākehā (New Zealand European) knowledge (Cunningham, 2000). This has led to the development of research methods and practices that support Māori practices and ways of knowing (Barnes, 2000; Powick, 2003; S. Walker et al., 2006). It is unclear whether data linkage as a research method is palatable for many potential participants of Māori ethnicity. Research on this issue may provide insight for researchers wishing to link survey data in New Zealand.

Building trust with participants is key to gaining consent. For Māori, that may mean building stronger relationships with Māori groups, ensuring that data is reported in a way that is respectful to Māori, and ensuring that research outcomes directly benefit Māori. The Health Research Council of New

Zealand has developed guidelines for health research involving Māori (Health Research Council of New Zealand, 2010). The guidelines acknowledge the important relationship between the Crown and Māori, based in the Tiriti o Waitangi / Treaty of Waitangi. The guidelines provide a consultation checklist, recommending that researchers build strong relationships with Māori groups, advisors, and cultural leaders from the outset of their project.

Overall health was not associated with consent

The second hypothesis of this study was that overall measures of health such as the SF-12 and overall health status would be associated with consent. Counter to the hypothesised relationship, in the current study overall measures of health were found not to be significantly related to consent in binary analysis ($p>0.05$) and had only small effect sizes. Overall measures were the SF-12 physical health score (d = -0.08) and the number of health conditions (ϕ = 0.05). Because their binary relationships with consent were not significant, these variables were excluded from the regression model.

This finding was at odds with many previous studies, which found overall health status indicators to be associated with consent to data linkage (Carter et al., 2010; Dunn et al., 2004; Klassen et al., 2005). It is notable that despite using several measures of overall health, the association with consent was not found in the current sample. Two explanations will be examined here. Firstly, there is a possibility that the SF-12 or SF-36 measure of health is not associated with consent. Secondly, and more importantly, the current study suggests that the overall health-consent association functions differently, or may not exist, for older adults.

The Short Form Health Survey may not be associated with consent. While the association between overall health and consent is apparent across studies, is not always found using a consistent or universal health measure. While other general health measures have been associated with consent to data linkage, the SF-12 has been shown not to be associated with consent in several samples. A survey of families in Hong Kong found that the SF-12 was not associated with consent to link survey information with administrative and medical datasets (Ni, Li, Hui, McDowell, & Leung, 2017).

In a nationally representative survey of mothers in the UK, Al Baghal (2015) found that the SF-12 was not significantly associated with consent to link either the mothers' or their children's data. In direct contradiction, another study of mothers in the US found that scores from the longer SF-36 were significantly related to consent (Woolf et al., 2009), however this was only in bivariate analysis. Huang, Shih, Chang, and Chou (2007) assessed the relationship between the SF-12 and consent, finding a significant relationship in bivariate analysis but not when other factors were accounted for

in multivariate analysis. This suggests that this specific health measure has no relationship with consent to linkage when other variables are accounted for. This does not rule out other measures of overall health being related to consent to link data.

The health-consent relationship may be different for older adults. Overall health conditions have been associated with consent to link across a variety of studies (Carter et al., 2010; Dunn et al., 2004; Klassen et al., 2005). In New Zealand, another longitudinal study known as SoFIE examined determinants of consent to link health data. The authors found a strong association between having two or more comorbid conditions and consent (Carter et al., 2010). This is contrary to the findings in the current study, where having more than one health condition was not associated with consent.

One key difference between the studies is the age of the groups surveyed. SoFIE included participants aged 15 and older, and other studies finding the health-consent relationship also used samples of varying ages, representative of the general population (Carter et al., 2010; Dunn et al., 2004; Klassen et al., 2005). In contrast HWR included only participants aged 60 and older when they were asked consent. This may suggest that health conditions have different effects for participants of different age groups. While poorer health may influence younger participants or those from a wide age range to consent, it may influence older adults differently.

As people age the likelihood that they have a health condition increases. Health concerns may become a routine part of life for older adults regardless of many of their circumstances (Moody & Sasser, 2015). In turn, this may mean that health concerns are normalised in older adult groups. They may in turn play less of a part in the decision making of older adults. This explanation would account for the lack of association between co-morbidity and consent decision.

The critical indication from this study is that, counter to studies in younger cohorts, overall health may not introduce consent bias in older adult samples where researchers wish to link data. As such researchers could approach health data linkage research with fewer concerns about bias, as such studies may be robust. Further research should be conducted to confirm the impact of overall health on consent in older adult samples. Specifically, researchers should investigate how the association may be mediated or moderated by age, salience, and other factors, and whether there is an underlying causal factor.

Diabetes was negatively associated with consent

While overall measures of health were not related to consent in the current study, interesting associations between specific illnesses and consent emerged. In bivariate analysis cancer and diabetes each had separate relationships with consent to linkage; the presence of cancer increased

the likelihood of data linkage while the presence of diabetes decreased it. Only the association between diabetes and consent remained significant in the multivariate analysis.

Specific health conditions have been variously associated with consent to linkage and so care must be taken to corroborate these associations, to ensure that they are not spurious or specific to the study. It is especially interesting that the diabetes-consent relationship holds while controlling for the Māori ethnicity-consent relationship in the same model. Older Māori have much higher incidence of diabetes than other ethnic groups in New Zealand (Ministry of Health, 2011). This suggests that each variable has an independent effect on consent.

One other study (Knies et al., 2012) also found diabetes to be related to consent. However, the relationship was reversed. Participants with diabetes were more likely to consent. There were several differences between the British Household Panel Survey (BHPS), which their finding is based on, and the HWR study. The BHPS began in 1992, with data collected through face-to-face interviews, the study focused on households rather than individuals, and people aged 16 or over were eligible to participate, making the study age range much wider.

If the consent decision for people with diabetes is related to capability, the difference in result could be explained by the difference in collection methods. Participants in the HWR completed the consent form alone and returned it in the post rather than having someone immediately present to collect their response. It is possible that this presented an additional barrier for people with diabetes who may have been less physically active or capable. However, if this were the case, other health conditions such as stroke and disability should also relate negatively to consent in the current study.

It is more likely that the authors' own explanation is the correct one – that people with diabetes were more likely to consent because diabetes was mentioned in the BHPS information leaflet (Knies et al., 2012). Participants may have seen diabetes as the 'symptom under investigation', a suggestion in line with findings by Dunn, Jordan, Lacey, Shapley, and Jinks (2004) that people consent at higher rates when they see themselves as having the symptom which is the focus of the study.

It is possible that this pattern of individuals consenting if they have the 'symptom under investigation' can be explained by the broader psychological construct of salience. This idea has been broached in other data linkage literature. In particular researchers have found that putting the consent question in context in the survey increases consent rates (Sala, Knies, & Burton, 2014). The authors suggest this works through increasing the salience of the question. Further theoretical work is required to understand the role of salience in consent to data linkage more fully.

Participating in more survey waves increased consent

The third hypothesis in this study was that participants who had participated for a greater number of waves would have higher rates of consent. The results supported this hypothesis in that the number of surveys that participants were included in was a key predictor of consent to link data (OR 1.88). This finding is congruent with the results of other studies (e.g. Mostafa, 2016; Sala et al., 2013). While little theory has addressed this area, greater participation in the survey could reflect a commitment to research, the development of rapport between the participant and the study, or greater familiarity.

Studies that have found an association between greater survey participation and consent to participate in research or link data use variables such as length of survey participation or number of surveys participated in. In his analysis Mostafa (2016) considered the number of times participants had responded to the survey as a proxy variable for 'loyalty to the survey'. He suggests his finding that participants who had responded to all waves consented at higher rates points to a "latent propensity to cooperate which underpins participation in the survey and participation in sub-studies" (Mostafa, 2016, p. 370). Further research is needed to establish an evidence base for this, or any other, hypothesis about this relationship.

What this finding does highlight is the importance of looking at variables related to the survey more broadly, such as 'para-survey' variables. Although they may not present the same apparent risk of biasing potential analyses, they are largely unresearched. Some survey variables do have the potential for researchers to control them. This enables researchers to maximise the possibility of gaining high rates of consent. For instance, some studies have found that certain interviewer characteristics (Korbmacher & Schroeder, 2013) or different placement of the linkage consent question (Sala et al., 2013) can have strong influences on consent. This is a key area of future research for those interested in increasing participation in linkage studies.

The regression model was a poor predictor of consent

While the results of this study were informative, and some supported previous research, the regression model in this study explained only a small amount of the variance in participants decision whether to consent to data linkage. Although there were systematic differences in consent by each of these variables, these differences were small and only accounted for 8-11% of a participant's consent decision overall. This is perhaps the most important finding from this study. It strongly suggests that older survey participants' decisions to consent to data linkage have relatively little to do with traditional health and wellbeing factors, and may instead be related to other factors.

It is unclear what these factors may be. Para-survey variables are a key place to start investigating. The role of the interviewer, placement of the consent question, prior participation, participant trust in the researcher and research generally, delivery mechanism for the survey, and how the consent decision is explained are examples of key variables of interest. It may be particularly important to investigate any interaction they may have with ethnicity, health, and other individual level, health related variables. Any interaction may also shed light on differences in findings between studies.

Overall, the results of the current study are positive for health researchers concerned about the degree of potential bias in analyses of health outcomes using this data. Because health-related factors only have a small impact on the consenting sample they are unlikely to introduce bias into any analysis. A variable must be associated with consent and both the independent and dependent variable under study to introduce consent bias to an analysis (Rothstein & Shoben, 2013). In this analysis it appears that consenters and non-consenters have relatively similar levels of health. So, while some factors do bias consent and may have an impact on study outcomes, that bias is likely to be small. This means that results from analyses with this data will be largely reliable.

The potential exception to this is diabetes. Researchers wishing to complete analyses of diabetes with the linked HWR data should take care to compensate for potential consent bias. Because diabetes was a significant predictor of consent, the proportions of participants with diabetes in the sample is likely to be skewed. More importantly if the research question uses diabetes as a predictor or outcome variable the risk of bias being introduced to the results is high.

A note of caution about study limitations

There are several limitations inherent to the current study. The study included a limited sample of ethnic minorities, was conducted in a specific context, and related only to health data. These should be considered when interpreting and generalising results.

The HWR sample includes an over sample of Māori participants, which provides adequate numbers for a statistical analysis of differences between Māori and New Zealand European participants. However, the survey attracted only small numbers of Pacific Peoples and people of other ethnicities. The small proportion of Pacific Peoples may have prevented the study from having sufficient power to find a significant relationship between Pacific ethnicity and consent.

The results of the current study may not be applicable to samples that are considerably different. It is unlikely to generalise to samples with a broader age distribution, younger samples, where consent to linkage is asked at the start of a longitudinal survey, or where linkage is completed outside of an academic research context. Further research should examine whether the results are generalisable.

Participants may consider health data fundamentally different from other kinds of data (Millican & Mansfield, 2013). They may have greater or lesser concerns about consenting to having their health data linked than other types of data such as financial or social data. As such, conclusions from this research may not apply to data linkage research with other types of data.

The future of consent in data linkage research

Findings like that of the current study identifying potential bias from data linkage consent have fueled calls from some researchers to soften or remove the requirements that researchers seek consent for data linkage (Breen, 2001; Holman, 2001; Ioannidis, 2013). Arguably, data linkage research poses minimal risk to individuals while providing the potential to contribute greatly to social good. However, data linkage is relatively new and has not often been discussed with the wider public (Jones, McNerney, & Ford, 2014). If any changes are to be made to consent requirements, and if the public is to have trust in how their data is used, the discussion must involve the wider public (e.g. Parkin & Paul, 2011).

Consent requirements were initially introduced to protect participants from being drawn into risky experiments against their will (Xafis, 2016). In New Zealand this was spurred by the Cartwright investigation into what became known as the 'unfortunate experiment', in which women with cervical cancer were unknowingly followed but not treated (Cartwright, 1988; Paul & Brookes, 2015). Informed consent in research became a legal requirement with the Health Research Council Act 1990 (O'Neill, 2013) and the Health Information Privacy Code outlines how health information must be treated (Privacy Commissioner, 1994). Given that these requirements were created in the context of experimental abuses, it is unclear if they are still fit for purpose with data linkage.

Some researchers argue that they are not. Often researchers cite the expense and impracticable nature of seeking consent from thousands of participants in large-scale research efforts (Breen, 2001; Holman, 2001; Singleton & Wadsworth, 2006) which may even prevent beneficial research from occurring (Xafis, 2016). Others argue that the risks of linkage are negligible. One author even suggests that consent seeking may be unnecessarily burdensome on participants (Singleton & Wadsworth, 2006). Eyal (2012) provides a clear overview of the arguments for and against consent in data linkage and Brownell and Jutte (2013) provide a compelling discussion of consent in relation to child abuse research.

New Zealand has unique issues to work through if researchers wish to increase focus on data linkage while maintaining respectful relationships with the indigenous Māori people. Māori information must be used in a respectful way (J. Walker et al., 2017). Māori are a collectivist people and health is

considered a community concern requiring community participation (Durie & Kingi, 1997). Data linkage increases the scale of the data that is made available to researchers dramatically. Māori see that group approval should be sought for the use of data (Massey University, 2013), so any change to requirements for accessing or using health information must involve active consultation with Māori.

Research into public perspectives on data linkage is scarce but can be a starting point to understand the issues associated with relaxing consent requirements. In one study parents agreed that linkage was important for vaccine surveillance and preferred minimal involvement with a consent process (Berry et al., 2013). Another study found that lay people found linkage without consent acceptable where the data was de-identified (Xafis, 2015). Across all studies, participants acknowledge the tension between individual rights and the collective good that can be achieved through linkage (Audrey, Brown, Campbell, Boyd, & Macleod, 2016; Berry et al., 2013; Xafis, 2015).

In New Zealand perspectives on sharing health information differ between Māori and Pākehā as well as within each group (Menkes, Hill, Horsfall, & Jaye, 2008). Individual autonomy is prized across groups, but Māori also see value in sharing health information with whanau (family groups). This suggests that New Zealanders are also capable of understanding and weighing the individual risks and collective benefits of data linkage.

If there is a need for consent requirements to be loosened or removed for data linkage studies, this discussion must be brought into the public domain. This is especially true in New Zealand, where indigenous Māori should be involved in decisions about Māori data. Additionally, the wider public must be informed about what is happening with their data, and its risks and benefits, for researchers to retain their trust. While some individuals may be averse to their data being linked, others are open to it. Research suggests that lay people are capable of understanding and weighing the issues of consent to linkage effectively.

Implications for health policy and practice

Data linkage has the potential to support the creation of data sets of a huge scale. The results of the current study suggest that, for older New Zealanders at least, the requirement for consent to data linkage is not likely to bias the result stemming from analysis of such linked data. This makes it an invaluable tool for understanding large scale health issues for older populations. In turn this research will support the development of health practice and policy, and go on to influence the wellbeing of large populations. If used well, this tool has potential to contribute greatly to population wellbeing.

Feder and Levitt (2005) outline three ways statistical analysis can play a role in health policy: defining a problem, developing a solution, and debating political action. The authors highlight the importance of providing the best quality analysis possible to reduce uncertainty in its conclusions and limit politicisation. When analyses are performed on data that is not representative of the population it brings the results into question, opening an avenue for those with a political agenda to question the truth of the findings (Feder & Levitt, 2005). This is a key contribution of analyses such as this one. It highlights the extent to which the linked data is biased by the consent process.

The move toward linked data sets in research is not without its challenges. Consent bias is an issue when linking survey data, as is the debate around removing consent requirements. Data must be stored and handled well, with effective controls in place for its use (Martin-Sanchez & Verspoor, 2014). Developing strong relationships with indigenous and minority groups is key to using their data respectfully and engendering trust, to ensure that they are accurately represented in the data (J. Walker et al., 2017).

Ultimately data linkage enables researchers to answer questions they otherwise couldn't. It allows researchers to find health trends, understand where to focus health spending (OECD, 2017), evaluate the influence of social variables on health outcomes, and much more. To enable data linkage to produce reliable results, researchers must understand the influence of consent requirements on the data they're working with. This is just as true for health research with older populations as it is for other groups.

References

Afshar, S., Roderick, P. J., Kowal, P., Dimitrov, B. D., & Hill, A. G. (2015). Multimorbidity and the inequalities of global ageing: a cross-sectional study of 28 countries using the World Health Surveys. *BMC Public Health*, *15*(1), 776. http://doi.org/10.1186/s12889-015-2008-7

Al-Shahi, R., Vousden, C., & Warlow, C. (2005). Bias from requiring explicit consent from all participants in observational research: prospective, population based study. *BMJ (Clinical Research Ed.)*, *331*(7522), 942. http://doi.org/10.1136/bmj.38624.397569.68

Al Baghal, T. (2015). Obtaining data linkage consent for children: factors influencing outcomes and potential biases. *International Journal of Social Research Methodology*, *5579*(July), 1–21. http://doi.org/10.1080/13645579.2015.1064635

Apfelbacher, C., Brandstetter, S., Herr, R., Ehrenstein, B., & Loerbroks, A. (2017). Depression and inflammatory arthritis are associated in both Western and Non-Western countries: Findings from the World Health Survey 2002. *Journal of Psychosomatic Research*, *92*, 49–54. http://doi.org/10.1016/j.jpsychores.2016.11.011

Audrey, S., Brown, L., Campbell, R., Boyd, A., & Macleod, J. (2016). Young people's views about consenting to data linkage: findings from the PEARL qualitative study. *BMC Medical Research Methodology*, *16*(1), 34. http://doi.org/10.1186/s12874-016-0132-4

Baker, R., Shiels, C., Stevenson, K., Fraser, R., & Stone, M. (2000). What proportion of patients refuse consent to data collection from their records for research purposes? *British Journal of General Practice*, *50*(457), 655–656.

Barnes, H. M. (2000). Kaupapa Maori: explaining the ordinary. *Pacific Health Dialogue*, *7*, 13–16.

Batra, G., Svennblad, B., Held, C., Jernberg, T., Johanson, P., Wallentin, L., & Oldgren, J. (2016). All types of atrial fi brillation in the setting of myocardial infarction are associated with impaired outcome, 1–8. http://doi.org/10.1136/heartjnl-2015-308678

Berry, J. G., Ryan, P., Duszynski, K. M., Braunack-Mayer, A. J., Carlson, J., Xafis, V., & Gold, M. S. (2013). Parent perspectives on consent for the linkage of data to evaluate vaccine safety: a randomised trial of opt-in and opt-out consent. *Clinical Trials*, *10*(3), 483–494. http://doi.org/10.1177/1740774513480568

Bishop, R. (1998). Freeing ourselves from neo-colonial domination in research: A maori approach to creating knowledge. *International Journal of Qualitative Studies in Education*, *11*(2), 119–219. http://doi.org/10.1080/095183998236674

Black, C., & Roos, L. L. (2005). Linking and combining data to develop statistics for understanding the population's health. In Daniel J Friedman, Edward L Hunter, & R Gibson Parrish II (Eds.), *Health statistics: Shaping policy and practice to improve the population's health* (pp. 214–240). New York: Oxford University Press.

Bo¨rsch-Supan, A., Hank, K., & Ju¨rges, H. (2005). A new comprehensive and international view on ageing: introducing the Survey of Health, Ageing and Retirement in Europe. *European Journal of Ageing*, *2*(4), 245–53.

Bohensky, M. a, Jolley, D., Sundararajan, V., Evans, S., Pilcher, D. V, Scott, I., & Brand, C. a. (2010). Data linkage: a powerful research tool with potential problems. *BMC Health Services Research*, *10*(1), 346. http://doi.org/10.1186/1472-6963-10-346

Borschmann, R., Thomas, E., Moran, P., Carroll, M., Heffernan, E., Spittal, M. J., … Kinner, S. A. (2016). Self-harm following release from prison: A prospective data linkage study. *Australian & New Zealand Journal of Psychiatry*. http://doi.org/10.1177/0004867416640090

Boyd, J. H., Randall, S. M., Ferrante, A. M., Bauer, J. K., Brown, A. P., & Semmens, J. B. (2014). Technical challenges of providing record linkage services for research. *BMC Medical Informatics and Decision Making*, *14*, 1–9. http://doi.org/10.1186/1472-6947-14-23

Bradley, C. J., Penberthy, L., Devers, K. J., & Holden, D. J. (2010). Health services research and data linkages: Issues, methods, and directions for the future. *Health Services Research*, *45*(5), 1468–1488. http://doi.org/10.1111/j.1475-6773.2010.01142.x

Breen, K. J. (2001). Consent for the linkage of data for public health research: is it (or should it be) an absolute pre-requisite? *Australian and New Zealand Journal of Public Health*, *25*(5), 423–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11688621

Brinkhof, M. W. G., Fekete, C., Chamberlain, J. D., Post, M. W. M., & Gemperli, A. (2016). Swiss national community survey of functioning after spinal cord injury: Protocol, characteristics of participants and determinants of non-response. *Journal of Rehabilitation Medicine*, *48*, 120–130.

Brownell, M. D., & Jutte, D. P. (2013). Administrative data linkage as a tool for child maltreatment research. *Child Abuse & Neglect*, *37*(2–3), 120–4. http://doi.org/10.1016/j.chiabu.2012.09.013

Buckley, B., Murphy, A. W., Byrne, M., & Glynn, L. (2007). Selection bias resulting from the requirement for prior consent in observational research: a community cohort of people with ischaemic heart disease. *Heart*, *93*(9), 1116–1120. http://doi.org/hrt.2006.111591 [pii]\r10.1136/hrt.2006.111591

Carter, K., Shaw, C., Hayward, M., & Blakely, T. (2010). Understanding the determinants of consent for linkage of administrative health data with a longitudinal survey. *Kotuitui: New Zealand Journal of Social Sciences Online*, *5*(2), 53–60. http://doi.org/10.1080/1177083X.2010.516440

Cartwright, S. R. (1988). *The Report of the Cervical Cancer Inquiry: The Report of the Committee of Inquiry into Allegations Concerning the Treatment of Cervical Cancer at NationalWomen's Hospital and Into Other Related Matters*. Wellington.

Cernin, P. A., Cresci, K., Jankowski, T. B., & Lichtenberg, P. A. (2010). Reliability and validity testing of the short-form health survey in a sample of community-dwelling African American older adults. *Journal of Nursing Measurement*, *18*(1), 49–59. http://doi.org/http://dx.doi.org/10.1891/1061-3749.18.1.49

Chamberlayne, R., Green, B., Barer, M. L., Hertzman, C., Lawrence, W. J., & Sheps, S. B. (1998). Creating a Population-based Linked Health Database: A New Resource for Health Services Research. *Canadian Journal of Public Health*, *89*(4), 270–273.

Charlton, R. A., Bettoli, V., Bos, H. J., Engeland, A., Garne, E., Gini, R., … Dolk, H. (2017). The limitations of some European healthcare databases for monitoring the effectiveness of pregnancy prevention programmes as risk minimisation measures. *European Journal of Clinical Pharmacology*. http://doi.org/10.1007/s00228-017-2398-9

Comino, E. J., Harris, M. F., Islam, M. D. F., Tran, D. T., Jalaludin, B., Jorm, L., … Haas, M. (2015). Impact of diabetes on hospital admission and length of stay among a general population aged 45 year or more: a record linkage study. *BMC Health Services Research*, *15*(1), 12. http://doi.org/10.1186/s12913-014-0666-2

Coolican, H. (2009). *Research Methods and Statistics in Psychology* (5th Editio). London: Hodder Education.

Cunningham, C. (2000). A framework for addressing Māori knowledge in research, science and technology. *Pacific Health Dialog*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11709883

da Silva, M. E. M., Coeli, C. M., Ventura, M., Palacios, M., Magnanini, M. M. F., Camargo, T. M. C. R., & Camargo, K. R. (2012). Informed consent for record linkage: a systematic review. *Journal of Medical Ethics*, *38*(10), 639–42. http://doi.org/10.1136/medethics-2011-100208

Dokholyan, R. S., Muhlbaier, L. H., & Falletta, J. M. (2009). Regulatory and ethical considerations for linking clinical and administrative databases. *American Heart Journal*, *157*(6), 971–982. http://doi.org/10.1016/j.ahj.2009.03.023

Doyle, M., & Sadler, K. (2013). *Welsh Health Survey: Study asking consent to link data*. Cardiff.

Dulin, P. L., Stephens, C., Alpass, F., Hill, R. D., & Stevenson, B. (2011). The impact of socio-contextual, physical and lifestyle variables on measures of physical and psychological wellbeing among Māori and non- Māori: the New Zealand Health , Work and Retirement Study The impact of socio-contextual , physical and lifestyle v. *Ageing and Society*, *31*(8), 1406–1424. http://doi.org/10.1017/S0144686X10001479

Dunn, K. M., Jordan, K., Lacey, R. J., Shapley, M., & Jinks, C. (2004). Patterns of consent in epidemiologic research: Evidence from over 25,000 responders. *American Journal of Epidemiology*, *159*(11), 1087–1094. http://doi.org/10.1093/aje/kwh141

Durie, M. H., & Kingi, T. K. R. (1997). *A framework for measuring Maori mental health outcomes, a research report prepared for the Ministry of Health (TPH 97/5)*. Palmerston North: Massey University, Department of Maori Studies.

Eyal, N. (2012). Informed Consent. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2012). Retrieved from http://plato.stanford.edu/archives/fall2012/entries/informed-consent/

Feder, J., & Levitt, L. (2005). Why truth matters: The role of health statistics in health policy. In D. J. Friedman, E. L. Hunter, & R. G. Parrish II (Eds.), *Health statistics: Shaping policy and practice to improve the population's health* (pp. 278–295). New York: Oxford University Press.

Fellegi, I. P., & Sunter, A. B. (1969). A Theory for Record Linkage.pdf. *Journal of the American Statistical Association*, *64*(328), 1183–1210.

Galenkamp, H., Deeg, D. J. H., de Jongh, R. T., Kardaun, J. W. P. F., & Huisman, M. (2016). Trend study on the association between hospital admissions and the health of Dutch older adults (1995-2009). *BMJ Open*, *6*(8), e011967. http://doi.org/10.1136/bmjopen-2016-011967

Glasson, E. J., & Hussain, R. (2008). Linked data: opportunities and challenges in disability research. *Journal of Intellectual & Developmental Disability*, *33*(4), 285–91. http://doi.org/10.1080/13668250802441409

Ha, N., Hendrie, D., & Moorin, R. (2014). Impact of population ageing on the costs of hospitalisations for cardiovascular disease: a population-based data linkage study. *BMC Health Services Research*, *14*(1), 554. http://doi.org/10.1186/s12913-014-0554-9

Harris, M. L., Dolja-Gore, X., Kendig, H., & Byles, J. E. (2016). End of life hospitalisations differ for older Australian women according to death trajectory: a longitudinal data linkage study. *BMC*

Health Services Research, *16*, 484. http://doi.org/10.1186/s12913-016-1729-3

Health Research Council of New Zealand. (2010). *Guidelines for Researchers on Health Research Involving Māori*. Auckland.

Heeringa, S. G., & Connor, J. H. (1995). *Technical Description of the Health and Retirement Survey Sample Design*. Ann Arbor, Michigan.

Holman, C. D. J. (2001). The impracticable nature of consent for research use of linked administrative health records. *Australian and New Zealand Journal of Public Health*, *25*(5), 421–2. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11688620

Holman, C. D. J., Bass, A. J., Rosman, D. L., Smith, M. B., Semmens, J. B., Glasson, E. J., … Stanley, F. J. (2008). A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system. *Australian Health Review*, *32*(4), 766–777.

Huang, N., Shih, S.-F., Chang, H.-Y., & Chou, Y.-J. (2007). Record linkage research and informed consent: who consents? *BMC Health Services Research*, *7*, 18. http://doi.org/10.1186/1472-6963-7-18

Ioannidis, J. P. A. (2013). Informed Consent, Big Data, and the Oxymoron of Research That Is Not Research. *The American Journal of Bioethics*, *13*(4), 40–42. http://doi.org/10.1080/15265161.2013.768864

Jensen, J., Spittal, M., Crichton, S., Sathiyandra, S., & Krishnan, V. (2002). The Direct Measurement of Living Standards: The New Zealand ELSI Scale. Nga Whakaaturanga Ahuatanga Noho.

Jensen, J., Spittal, M., & Krishnan, V. (2005). *ELSI Short Form: User Manual for a Direct Measure of Living Standards*. Wellington: Ministry of Social Development.

Jones, K. H., McNerney, C. L., & Ford, D. V. (2014). Involving consumers in the work of a data linkage research unit. *International Journal of Consumer Studies*, *38*, 45–51. http://doi.org/10.1111/ijcs.12062

Jutte, D. P., Roos, L. L., & Brownell, M. D. (2011). Administrative record linkage as a tool for public health research. *Annual Review of Public Health*, *32*, 91–108. http://doi.org/10.1146/annurev-publhealth-031210-100700

Kent, E. E., Ambs, A., Mitchell, S. A., Clauser, S. B., Smith, A. W., & Hays, R. D. (2015). Health-related quality of life in older adult survivors of selected cancers: Data from the SEER-MHOS linkage. *Cancer*, *121*(5), 758–765. http://doi.org/10.1002/cncr.29119

Kho, M. E., Duffett, M., Willison, D. J., Cook, D. J., & Brouwers, M. C. (2009). Written informed consent and selection bias in observational studies using medical records: systematic review. *Bmj*, *338*(b866), 822. http://doi.org/10.1136/bmj.b866

Klassen, A. F., Lee, S. K., Barer, M., & Raina, P. (2005). Linking survey data with administrative health information: Characteristics associated with consent from a neonatal intensive care unit follow-up study. *Canadian Journal of Public Health*, *96*(2), 151–154.

Knies, G., Burton, J., & Sala, E. (2012). Consenting to health record linkage: evidence from a multi-purpose longitudinal survey of a general population. *BMC Health Services Research*, *12*(1), 52. http://doi.org/10.1186/1472-6963-12-52

Korbmacher, J. M., & Schroeder, M. (2013). Consent when Linking Survey Data with Administrative

Records: The Role of the Interviewer. *Survey Research Methods*, *7*(2), 115–131. Retrieved from https://ojs.ub.uni-konstanz.de/srm/article/view/5067

Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P. ., Helenius, H., & Koskenvuo, M. (2001). Non-Response and Related Factors in a Nation-Wide Health Survey. *European Journal of Epidemiology*, *17*(11), 991–999.

Lipowski, E. E., & Bigelow, W. E. (1996). Data Linkages for Research on Outcomes of Long-Term Care. *The Gerontologist*, *36*(4), 441–447.

Lombardo, P. A. (2006). Eugenics, Medical Education, and the Public Health Service: Another Perspective on the Tuskegee Syphilis Experiment. *Bulletin of the History of Medicine*, *80*(2), 291–316. http://doi.org/10.1353/bhm.2006.0066

Macpherson, L. (2015). *New Zealand period life tables: 2012–14 key facts*. Statistics New Zealand.

Madans, J. H., & Cohen, S. B. (2005). Health Surveys: A Resource to Inform Health Policy and Practice. In D. J. Friedman, E. L. Hunter, & G. R. Parish II (Eds.), *Health statistics: Shaping policy and practice to improve the population's health* (pp. 119–138). New York: Oxford University Press.

Maguire, A., & O'Reilly, D. (2015). Does conurbation affect the risk of poor mental health? A population based record linkage study. *Health & Place*, *34*, 126–34. http://doi.org/10.1016/j.healthplace.2015.05.003

Martin-Sanchez, F., & Verspoor, K. (2014). Big Data in Medicine Is Driving Big Changes. *IMIA Yearbook*, *9*(1), 14–20. http://doi.org/10.15265/IY-2014-0020

Massey University. (2013). Code of ethical conduct for research, teaching and evaluations involving human participants. Author. Retrieved from http://www.massey.ac.nz/massey/research/ethics/human-ethics/ code/code.cfm

Melton, L. J. (1996). History of the Rochester Epidemiology Project. *Mayo Clinic Proceedings. Mayo Clinic*, *71*(3), 266–274. http://doi.org/10.4065/71.3.266

Menkes, D. B., Hill, C. J., Horsfall, M., & Jaye, C. (2008). Perspectives on access to personal health information in New Zealand / Aotearoa. *Anthropology and Medicine*, *15*(3), 199–212. http://doi.org/10.1080/13648470802355608

Millican, C., & Mansfield, C. (2013). *Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data*. London. Retrieved from https://wellcome.ac.uk/sites/default/files/wtp053205_0.pdf

Ministry of Health. (2011). *Tatau Kura Tangata: Health of Older Maori Chart Book 2011*. Wellington: Ministry of Health. Retrieved from http://www.moh.govt.nz

Ministry of Health. (2015). Tatau Kahukura: Māori health statistics. Retrieved June 22, 2017, from http://www.health.govt.nz/our-work/populations/maori-health/tatau-kahukura-maori-health-statistics

Ministry of Health. (2016). *Annual Update of Key Results 2015/16: New Zealand Health Survey*. *New Zealand Health Survey. Wellington: Ministry of Health*. Wellington. Retrieved from http://www.health.govt.nz/system/files/documents/publications/annual-update-key-results-2015-16-nzhs-dec16-v2.pdf

Moody, H. R., & Sasser, J. R. (2015). *Aging: Concepts and Controversies* (8th ed.). Los Angeles: SAGE Publications.

Morton, S. M. B., Atatoa Carr, P. E., Bandara, D. K., Grant, C. C., Ivory, V. C., Kingi, T. R., … Waldie, K. E. (2010). *Growing Up in New Zealand: A longitudinal study of New Zealand children and their families. Report 1: Before we are born.* Auckland: Growing Up in New Zealand.

Mostafa, T. (2016). Variation within households in consent to link survey data to administrative records: evidence from the UK Millennium Cohort Study. *International Journal of Social Research Methodology*, *19*(3), 355–375. http://doi.org/10.1080/13645579.2015.1019264

Ni, M. Y., Li, T. K., Hui, R. W. H., McDowell, I., & Leung, G. M. (2017). Requesting a unique personal identifier or providing a souvenir incentive did not affect overall consent to health record linkage: evidence from an RCT nested within a cohort. *Journal of Clinical Epidemiology*, *84*, 142–149. http://doi.org/10.1016/j.jclinepi.2017.01.003

O'Neill, J. (2013). Creative Research Ethics in the Enterprise University: what price academic freedom? *Policy Futures in Education*, *11*(6), 712. http://doi.org/10.2304/pfie.2013.11.6.712

OECD. (2017). *Tackling wasteful spending on health*. Paris: OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264266414-en

Pagano-Therrien, J. (2013). Exploring Research Fatigue in HIV-Infected Youth. *Journal of the Association of Nurses in AIDS Care*, *24*(1), 11–16. http://doi.org/10.1016/j.jana.2012.01.010

Park, J. H., Lim, S., Lim, J., Kim, K.-I., Han, M.-K., Yoon, I. Y., … Kim, K. W. (2007). An Overview of the Korean Longitudinal Study on Health and Aging. *Psychiatry Investigation*, *4*, 84–95.

Parkin, L., & Paul, C. (2011). Public good, personal privacy: a citizens' deliberation about using medical information for pharmacoepidemiological research. *Journal of Epidemiology and Community Health*, *65*(2), 150–156. http://doi.org/10.1136/jech.2011.142976a.75

Paul, C., & Brookes, B. (2015). The rationalization of unethical research: Revisionist accounts of the Tuskegee syphilis study and the New Zealand "unfortunate experiment." *American Journal of Public Health*, *105*(10), e12–e19. http://doi.org/10.2105/AJPH.2015.302720

Pickrell, W. O., Lacey, A. S., Bodger, O. G., Demmler, J. C., Thomas, R. H., Lyons, R. A., … Kerr, M. P. (2015). Epilepsy and deprivation, a data linkage study. *Epilepsia*, *56*(4), 585–591. http://doi.org/10.1111/epi.12942

Powick, K. (2003). *Maori research ethics*. Hamilton: Wilf Malcolm Institute of Educational Research, University of Waikato.

Privacy Commissioner. (1994). Health Information Privacy Code. Auckland: KB Print. Retrieved from www.privacy.org.nz

Regidor, E. (2004). The use of personal data from medical records and biological materials: Ethical perspectives and the basis for legal restrictions in health research. *Social Science and Medicine*, *59*, 1975–1984. http://doi.org/10.1016/j.socscimed.2004.02.032

Resnick, B., & Nahm, E. S. (2001). Reliability and validity testing of the revised 12-item short form health survey in older adults. *Journal of Nursing Measurement*, *9*(2), 151–161.

Roos, L. L., Brownell, M., Lix, L., Roos, N. P., Walld, R., & MacWilliam, L. (2008). From health research to social research: Privacy, methods, approaches. *Social Science & Medicine*, *66*, 117–129.

http://doi.org/10.1016/j.socscimed.2007.08.017

Rothstein, M. A., & Shoben, A. B. (2013). Does Consent Bias Research? *The American Journal of Bioethics*, *13*(4), 27–37. http://doi.org/10.1080/15265161.2013.767955

Sakshaug, J., Couper, M. P., Ofstedal, M. B., & Weir, D. R. (2016). Linking survey and administrative records: Mechanisms of consent. *Sociological Methods & Research*, *41*(4), 535–569. http://doi.org/10.1177/0049124112460381.LINKING

Sakshaug, J., & Kreuter, F. (2012). Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data. *Survey Research Methods*, *6*(2), 113–122.

Sala, E., Burton, J., & Knies, G. (2012). Correlates of Obtaining Informed Consent to Data Linkage: Respondent, Interview, and Interviewer Characteristics. *Sociological Methods & Research*, *41*(3), 414–439. http://doi.org/10.1177/0049124112457330

Sala, E., Knies, G., & Burton, J. (2013). *Propensity to Consent to Data Linkage: Experimental Evidence from the Innovation Panel on the Role of Three Survey Design Features* (Understanding Society Working Paper Series No. 2013–5). Colchester, Essex.

Sala, E., Knies, G., & Burton, J. (2014). Propensity to consent to data linkage: Experimental evidence on the role of three survey design features in a UK longitudinal panel. *International Journal of Social Research Methodology*, *17*, 455–473.

Salmond, C., Crampton, P., & Atkinson, J. (2007a). *NZDep2006 Index of Deprivation*. *Department of Public Health*. Wellington: Department of Public Health, University of Otago. Retrieved from http://www.otago.ac.nz/wellington/otago020337.pdf

Salmond, C., Crampton, P., & Atkinson, J. (2007b). *NZDep2006 Index of Deprivation user's manual*. *Department of Public Health*. Wellington: Department of Public Health, University of Otago. Retrieved from http://www.otago.ac.nz/wellington/otago020337.pdf

Sharp, L. M., & Frankel, J. (1983). Respondent Burden: A Test of Some Common Assumptions. *Public Opinion Quarterly*, *47*(1), 36–53. http://doi.org/10.1086/268765

Singleton, P., & Wadsworth, M. (2006). Consent for the use of personal medical data in research. *BMJ (Clinical Research Ed.)*, *333*(7561), 255–8. http://doi.org/10.1136/bmj.333.7561.255

Statistics New Zealand. (2004). *Statistical Standard for ethnicity*. *Report of the Review of the Measurement of Ethnicity*. Wellington.

Statistics New Zealand. (2012). *Privacy impact assessment for the Integrated Data Infrastructure*. Wellington. Retrieved from www.stats.govt.nz

Stewart, R., & Davis, K. (2016). "Big data" in mental health research: current status and emerging possibilities. *Social Psychiatry and Psychiatric Epidemiology*, *51*(8), 1055–1072. http://doi.org/10.1007/s00127-016-1266-8

Sukumar, D. W., Harvey, L. A., Mitchell, R. J., & Close, J. C. T. (2016). The impact of geographical location on trends in hospitalisation rates and outcomes for fall-related injuries in older people. *Australian and New Zealand Journal of Public Health*, *40*(4), 342–348. http://doi.org/10.1111/1753-6405.12524

Taylor, R., Conway, L., Calderwood, L., Lessof, C., Cheshire, H., Cox, K., & Scholes, S. (2007). *Health, Wealth and Lifestyles of the Older Population in England: The 2002 English Longitudinal Study*

of Ageing. London. Retrieved from http://www.ifs.org.uk/elsa/report03/w1_tech.pdf

Todd, A. L., Aitken, C. A., Boyd, J., & Porter, M. (2016). Testing a health research instrument to develop a state-wide survey on maternity care, *26*(January), 1–5.

Towers, A., Stevenson, B., Breheny, M., & Allen, J. (2017). Health, Work, and Retirement Longitudinal Study. In N. Pachana (Ed.), *Encyclopedia of Geropsychology* (pp. 1025–1033). Singapore: Springer Singapore. http://doi.org/10.1007/978-981-287-082-7

United Nations Department of Economic and Social Affairs Population Division. (2015). *World population ageing*. New York. Retrieved from http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf%5Cnwww.un.org/.../population/.../WPA2009/WPA2009

Vasunilashorn, S., Glei, D. A., Weinstein, M., & Goldman, N. (2013). Perceived stress and mortality in a Taiwanese older adult population. *Stress (Amsterdam, Netherlands)*, *16*(6), 600–6. http://doi.org/10.3109/10253890.2013.823943

Walker, J., Lovett, R., Kukutai, T., Jones, C., & Henry, D. (2017). Indigenous health data and the path to healing. *The Lancet*, *390*(10107), 2022–2023. http://doi.org/10.1016/S0140-6736(17)32755-1

Walker, S., Eketone, A., & Gibbs, A. (2006). An exploration of kaupapa Maori research, its principles, processes and applications. *International Journal of Social Research Methodology*, *9*(4), 331–344. http://doi.org/10.1080/13645570600916049

Walter, L. C., Bertenthal, D., Lindquist, K., & Konety, B. R. (2006). PSA screening among elderly men with limited life expectancies. *Jama*, *296*(19), 2336–2342. http://doi.org/296/19/2336 [pii]\r10.1001/jama.296.19.2336

Walter, L. C., Lindquist, K., Nugent, S., Schult, T., Lee, S. J., Casadei, M. A., & Partin, R. (2009). Impact of age and comorbidity on colorectal cancer screening among older veterans. *Annals of Internal Medicine*, *150*(7), 465–473. http://doi.org/10.1523/JNEUROSCI.3593-07.2007.Omega-3

Ware, J., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey : Construction of Scales and Preliminary Tests of Reliability and Validity Author (s): John E . Ware , Jr ., Mark Kosinski and Susan D . Keller Published by : Lippincott Williams & Wilkins Stable URL : http://www.jstor, *34*(3), 220–233.

Winkler, W. E. (1999). The State of Record Linkage and Current Research Problems. *Statistical Research Division US Census Bureau*, 1–15. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.39.4336

Woolf, S. H., Rothemich, S. F., Johnson, R. E., & Marsland, D. W. (2009). Selection bias from requiring patients to give consent to examine data for health services research. *Archives of Family Medicine*, *9*(10), 1111–1118. http://doi.org/10.1001/archfami.9.10.1111

World Health Organization. (2015). *World report on ageing and health*. Luxembourg: WHO Press.

World Health Organization. (2017). WHO World Health Survey. Retrieved from http://www.who.int/healthinfo/survey/en/

Xafis, V. (2015). The acceptability of conducting data linkage research without obtaining consent: lay people's views and justifications. *BMC Medical Ethics*, *16*(79), 1–16. http://doi.org/10.1186/s12910-015-0070-4

Xafis, V. (2016). What consent model is ethically justifiable in cancer population research? *Ethics in Cancer*, *40*(2), 102–106.

Yawn, B. P., Yawn, R. A., Geier, G. R., Xia, Z., & Jacobsen, S. J. (1998). The impact of requiring patient authorization for use of data in medical records research. *Journal of Family Practice*, *47*(5), 361–365.

Young, A. F., Dobson, A. J., & Byles, J. E. (2001). Health services research using linked records: who consents and what is the gain? *Australian and New Zealand Journal of Public Health*, *25*(5), 417–420. http://doi.org/10.1111/j.1467-842X.2001.tb00649.x