




The value of administrative data for longitudinal social research: a case study investigating income support receipt and relationship separation in Australia

Melanie Spallek, Michele Haynes, Janeen Baxter & Nicole Kapelle


To cite this article: Melanie Spallek, Michele Haynes, Janeen Baxter & Nicole Kapelle (2020): The value of administrative data for longitudinal social research: a case study investigating income support receipt and relationship separation in Australia, International Journal of Social Research Methodology, DOI: [10.1080/13645579.2019.1707984](https://doi.org/10.1080/13645579.2019.1707984)

To link to this article: <https://doi.org/10.1080/13645579.2019.1707984>

 View supplementary material 

 Published online: 08 Jan 2020.





 Submit your article to this journal 

 View related articles 

 View Crossmark data 



The value of administrative data for longitudinal social research: a case study investigating income support receipt and relationship separation in Australia

Melanie Spallek ^a, Michele Haynes ^a, Janeen Baxter ^b and Nicole Kapelle ^{b,c}

^aInstitute for Learning Sciences and Teacher Education, Australian Catholic University, Brisbane, Australia; ^bARC Centre of Excellence for Children and Families over the Life Course, Institute for Social Science Research, The University of Queensland, Brisbane, Australia; ^cDepartment of Social Sciences, Humboldt-Universität zu Berlin, Berlin, Germany

ABSTRACT

In recent years, the Australian government has encouraged open access to administrative data, providing new opportunities for examining life course pathways and evaluating social policies. This paper demonstrates the importance of establishing partnerships in the use of administrative data for social research. In collaboration with the data custodian, we used administrative welfare data to investigate a policy-relevant topic: the association between income support receipt and relationship separation. Our results provided greater statistical power for detecting associations for minority groups than is possible with panel survey data, highlighting the benefits of administrative data for understanding the outcomes of population groups that are not well represented in surveys. The collaboration between university researchers and data custodians was critical to enable appropriate sharing of data for research and for accurate interpretation of the data and outcomes for relevance to policy.



KEYWORDS


Administrative data;
longitudinal data analysis;
event-history analysis;
income support payments;
relationship separation

Introduction

Research that addresses transitions and change in social and economic outcomes for individuals has often relied on national panel surveys that collect data from the same cohort or population sample at regular intervals over a long period of time. The availability of data from high-quality longitudinal surveys is invaluable for informing social research on life course pathways and outcomes. This type of survey data, however, is not always appropriate due to potential distortion of measures, insufficient sample sizes for minority groups or design issues related to, for instance, the timing and interval of interviews. Administrative data are an alternative source for research and refers to data that has been collected from the operation of administrative systems for services provided by the government or community agencies, for example, the provision of income support payments to eligible individuals and families (Elias, 2014). These data are a useful source of information for social research, providing new opportunities to examine social behaviour, life course pathways, and evaluations of social policies and programs (Connelly, Playford, Gayle, & Dibben, 2016; Crichton, Templeton, & Tumen, 2015).

One of the major strengths of administrative data is the coverage of hard-to-reach populations (e.g. single mothers, ethnic minority groups) particularly for social stratification research that

CONTACT Melanie Spallek  Melanie.Spallek@acu.edu.au  Institute for Learning Sciences and Teacher Education, Australian Catholic University, Brisbane Campus, Level 4, 229 Elizabeth Street, Brisbane CBD, QLD 4000, Australia

 Supplemental data for this article can be accessed [here](#).

focuses on the causes and consequences of social disadvantage. Another considerable advantage is the accurate measurement of the information related to the timing of provision of a service to an individual as well as information about the characteristics of the individual and other family members. Whereas panel data is often only collected in larger intervals and relies on respondents to accurately recall detailed information such as dates and timing of events, administrative data is collected for the purpose of delivering a service and precisely measures the start and end time of an event of interest. Although administrative data may be limited in some ways (e.g. scope of measures) and may involve extensive data management procedures to prepare the data for analysis, they can be an alternative and complementary source for objective measures and new information on the provision of services for the population of interest that would be otherwise unavailable for research (Connelly et al., 2016).

The provision of some form of open access to public sector data for research purposes has had long traditions in countries such as the Norway, Sweden and The Netherlands. In Australia, however, access to government administrative data for research has been restricted, although there are some variations in the extent of restrictions across States and Territories. In 2017 an inquiry by the Australian Productivity Commission recommended more open access to administrative data (Productivity Commission, 2017). This recommendation was in part informed by best-practice examples from countries such as New Zealand where the social investment approach has promoted the potential of government administrative data to inform efforts to reduce welfare burdens through targeted investment to support at-risk groups. Australian social policy has moved in similar directions under the Commonwealth Government's Priority Investment Approach to Welfare which has used actuarial analyses of data collected for the delivery of social security and welfare payments to identify cohorts at risk on long-term welfare dependence.

This paper outlines a collaboration with the Australian Government Department of Human Services (DHS) that was formed to trial a methodology that would enable a research team to analyse a full population of Commonwealth Government administrative data to address a question of mutual interest to both Government Policy and researchers. The methodological challenges to accessing and analysing government administrative data are considerable. These include privacy and security requirements for analysis of the data, limitations to the scope of data, and inconsistencies in the recording of the data over time. Importantly, the use of administrative data requires establishing a respectful partnership between researchers and data custodians for an agreed purpose.

First, we introduce the case study and the background for the research question concerning the association between income support receipt and relationship breakdown. Second, we describe the methods used to overcome the restrictions for non-government personnel to access the data and outline the methodology for the analysis given the restrictions to access and limitations of the data. Third, we illustrate the use of administrative data for research by analysing the association between the probability of remaining in receipt of income support following a relationship breakdown and the methodological considerations for the application of an event-history model. We assess the potential of administrative data for this research by comparing the feasibility of a similar approach as applied to panel data obtained from the Households, Income and Labour Dynamics in Australia survey (HILDA), a high quality, large-scale longitudinal public use survey which also collects information on income support and relationship breakdown. In doing so, we provide insight into both substantive questions about relationship breakdown and length of time on government income support, as well as the analytical strengths and limitations of administrative data.

Fourth, we discuss the advantages and limitations of this approach including the limitations to building a statistical model that captures the complexity of the service provision processes. In doing so, we provide insight into the challenges and benefits of administrative data for longitudinal social research and policy analysis.

This collaboration offered an exciting new opportunity to investigate the methodological issues associated with using a unique source of population data that had not previously been available for

research or policy analysis. This study was one of the first in Australia to analyse an entire population of cases who have accessed certain government income support payment services, requiring the development of a close partnership and clear protocols for accessing, analysing and reporting of the data.

The case study: income support receipt and relationship separation

Why use administrative data?

The use of survey data to examine the length of time in receipt of income support has been shown to be relatively unreliable due to the limitations and biases of memory and recall (Blank & Ruggles, 1994; Pavetti, 1994). More reliable results regarding the length of time in receipt of income support may be sourced from administrative data (Barrett, 2000; Blank, 1989; Wilson, 1999). Compared to survey data, which draws a representative sample from the population, administrative data includes each individual and family of the population that receive the service. This improves the accuracy of the inference from the data available. Additionally, administrative data also facilitates examination of minority populations which are frequently under-represented in survey samples (Connelly et al., 2016). While government administrative datasets may provide complete information on individuals' income support payments, they do not typically record information on individuals during periods that they are not eligible for payments. It is therefore not possible to observe the complete income trajectories of individuals who no longer receive income support. Government administrative datasets are also often deficient in measures that are important for social research and may present inconsistencies in reporting of measures over time, as the data is not collected for research purposes (Connelly et al., 2016). Generally, these datasets are not accompanied by documentation, which poses a further practical challenge for researchers wanting to utilise these data sources.

In Australia, a limited number of studies have previously investigated receipt of income support payments using a subset of the DHS administrative data called the Longitudinal Data Set (LDS) (Tseng, Vu, & Wilkins, 2008; Tseng & Wilkins, 2003), which contained fortnightly income support payment records for a one percent sample of income recipients over a five-and-a-half-year data window from January 1995 to June 2001. This body of research reported that although a significant number of individuals relied on income support payments temporarily, a large number became reliant long-term (Tseng & Wilkins, 2003) which was associated with the Age Pension payment.

Determinants of income support reliance

A range of factors are associated with entry to and exit from income support payments (Stellmack, Wanberg, & Kammeyer-mueller, 2003). These include education, work experience and occupational skills, the presence and age of children, ethnicity and English language skills. While higher levels of education have been found to be positively associated with employment and less reliance on income support (Bora, Caudill, Spera, & Kunz, 1998; Harris, 1993; Kroch & Sjoblom, 1994; Meyer & Cancian, 1998), having limited education has been linked to longer periods on income support payments (Bane & Ellwood, 1983; Coe, 1981; Petersen, 1995). Prior work experience and occupational skills have also been reported to enable employment (Cheng, 2002; Leahy, Buss, & Quane, 1995). The presence of children (especially young) has a disadvantageous impact on exiting from income support payments, in particular when there is a need for childcare (Stellmack et al., 2003). Ethnic minorities have been found to be less likely to leave the welfare system (Bane & Ellwood, 1983; Cheng, 2002; Gault, Hartmann, & Yi, 1998; Meyers & Heintze, 1999; Piskulich, 1993) but those with higher proficiency with the English language were shown to have higher probabilities of moving off income support (Stellmack et al., 2003). In Australia, demographic characteristics shown to be associated with a long-term reliance on income support payments also include Indigenous and refugee status (Department of Social Services, 2017).

Income support receipt and relationship separation

One of the main determinants of income support receipt for women is relationship breakdown. Studies using survey data have demonstrated an association between separation from a relationship and decreased economic wellbeing, including evidence that these patterns are gendered with women faring worse than men, leading to prolonged financial instability and long-term support dependence (Bayaz-Ozturk, Burkhauser, Couch, & Hauser, 2018; De Vaus, Gray, Qu, & Stanton, 2014; Maroto, 2015). A combination of social and economic factors is likely to explain why women experience greater financial hardship following relationship breakdown compared to men. Although women's participation in the labour market has increased markedly since the 1970s, women in Western societies, including Australia, undertake the primary care of children, and many partnered women typically withdraw from the labour market or reduce their hours of employment when they have children (Cipollone, Patacchini, & Vallanti, 2014; Craig, Mullan, & Blaxland, 2010; Gaudet, Cooke, & Jacob, 2011; Pearce, 2000). This gendered division of care work means that women who experience relationship separation are more at risk of financial hardship than men due to the loss of support from partner earnings and more likely to be in receipt of welfare support (De Vaus et al., 2014; De Vaus, Gray, Qu, & Stanton, 2015; Sheehan, 2002; Bruce Smyth, Rodgers, Son, & Vnuk, 2015; B Smyth & Weston, 2000). Women who wish to re-enter the labour market after relationship breakdown may not be able to do so because of the loss of work experience and career interruption from care responsibilities (Tamborini, Couch, & Reznik, 2015; Van Damme, 2010).

Findings for men are mixed, but typically show a less detrimental effect of relationship breakdown on men's household income (Andreß, Borgloh, Bröckel, Giesselmann, & Hummelsheim, 2006; De Vaus et al., 2014; McKeever & Wolfinger, 2001; Tach & Eads, 2015). This may be due to the fact that marriage and parenthood are typically associated with men's connection to the labour market and earnings. Hence, their earnings capacity after separation is typically greater compared to women (Andreß et al., 2006; De Vaus et al., 2014).

In this case study, we take a longitudinal approach to address the question on associations between relationship separation and income support payments reliance and make use of detailed fortnightly information on income support payments ranging over a ten-year period from 2003 to 2013. Using administrative data, we have the advantages of a large population cohort and coverage of a hard-to-reach population that is typically under-represented in survey collections, as well as high reliability of measures related to receipt of income support, as the data is collected for the purposes of delivery of this payment. These data were extracted from an existing database and therefore we were not faced with the considerable challenges associated with analysing linked data (Harron et al., 2017).

We include both men and women and examine the impact of relationship separation on the likelihood of remaining on income support payments. We limit the age group of the analytical sample so that Age Pension payments are excluded, since men and women become eligible for Age Pension at the age of 65 years with some residence requirements (Department of Social Services, Department of Agriculture and Water Resources, Department of Employment, & Department of Education and Training, 2017). Including Age Pension recipients in our analysis would lead to biased estimates of the impact of separation on income support payments reliance.

In the next section, we outline the methodology required to effectively partner with the government agency to access the data and the approach to analysing and generating new evidence from their administrative data.

Research methodology

Approach to collaboration with the data custodian

DHS is the data custodian of the highly sensitive data on income support recipients in Australia and therefore has administrative responsibilities for the data. To undertake research using the relevant

administrative data, commitment to the project and a close collaboration between Department personnel and the academic research team was essential. Due to security restrictions and legislations related to privacy and confidentiality associated with the administrative data, it was agreed that a research fellow from the Australian Research Council Centre of Excellence for Children and Families over the Life Course (Life Course Centre) would be provided a placement position at the data custodian's office to analyse the data in a secure environment. The data custodian's personnel assigned to this project had extensive experience with the administrative database and its contents, as well as expertise and knowledge of the systems and payments. In collaboration with the data custodian, the population of interest and variables required to conduct the analysis were identified and extracted from the database by the data custodian's personnel. This data is not publicly available, and researchers interested in accessing this data for research purposes would need to go through similar processes.

Challenges for a collaborative approach included the requirement for the physical location of the researcher on the government site, which hindered discussions on analysis decision amongst the research team. This was further delayed by the requirement for all output produced to be checked and verified by the data custodian's personnel before it could be taken outside the data custodian's office. The statistical analysis software used was restricted to that endorsed within the Department, in this case SAS[®] analytics software (SAS Institute Inc., 2013).

Extraction of administrative data on income support payments for the population of interest

Decisions on the extent to which the data could be disaggregated for analysis and reporting were made in consultation with the data custodian's personnel. The administrative data contained records on more than eight million people who were actively in receipt of income support payments. This information was in addition to the millions of historical records for people who were no longer receiving income support payments. To address the case study research question on the association between duration of income support and relationship status, the analytical sample was restricted to individuals aged 15–54 years, eligible for any type of income support payment on the 30th of June 2003 and who had received payments the previous financial year, between 1 July 2002 and 30 June 2003. For a complete list of income support payments and eligibility criteria refer to the Australian Government Department of Human Services website (<https://www.humanservices.gov.au/individuals/topics/income-support-payment-description/34696>). Further, the sample was restricted to those registered as partnered on the 30 June 2003. This resulted in a dataset containing 538,365 individuals. As noted above, restricting the sample to those aged 15–54 years old in 2003 ensured that we mostly excluded individuals on Age Pension or starting on Age Pension before the end of the observation period which, for the purpose of this research, was defined as the 30 June 2013.

The final number of individuals included in the analysis was 537,912 and excluded the following those who received an income support payment for 1 day only, which usually referred to one-off payments or lump sum payments. Similarly, to those eligible for Age Pension, individuals receiving one-off payments were not relevant to this analysis as separation would not have an impact on the time spent receiving income support.

Measures for analysis

Time at risk: receiving income support payment

The Department collects daily information on the receipt and type of payment allocated to an individual. This information is updated fortnightly in the database. A period on income support was defined from the first date of receiving any income support to the last date of receiving a payment. However, if an individual temporarily exited income support and returned to receiving support within three months of the last day of the previous period, the two periods

were combined and defined as one period of income support. One period on income support may include a combination of several different payment types. For example, an individual's entitlement could change from Parental Payment Partnered to Parental Payment Single. As we were interested in receipt of income support rather than the type of support received, these changes in payment type were not taken into consideration for these analyses, with the exception for those on Disability Support Pensions (DSP). Through discussions with the DHS personnel, we learned that this was necessary as disability payments are typically associated with irreversible long-term conditions. An individual is very unlikely to have payments of this type discontinued following confirmation of eligibility. Failing to control for DSP would have resulted in an underestimation of the effect of separation on income support reliance. Our analyses therefore controlled for whether the individual had received DSP during the period in receipt of income support, coded as [1] yes, [0] no.

Relationship between separation and demographic variables

Information on relationship status is collected while a person receives income support and the start date and end date of the relationship status is collected. There is generally no information on relationship status available for individuals once they have exited the payment system. The information captured with each variable depends primarily on the type of information needed for administrative purposes. With regards to relationship status, the extent of information needed for the Department to assess income support eligibility is whether a person is partnered or not and does not distinguish relationship type, such as cohabiting, married, divorced or separated. For this reason, the value of relationship status was defined as partnered [1] or not partnered [0].

As noted above, previous research has shown that certain demographic characteristics are associated with income support. The data provided information on the following characteristics which were included as time-constant covariates in the analysis: Gender coded as [1] male and [0] female; age in 2003 grouped into [1] 15–24, [2] 25–34, [3] 35–44, [4] 45–54; Indigenous status coded as [1] Indigenous, [0] non-Indigenous; refugee indicator in 2003 coded as [1] yes, [0] no; and non-English speaking indicator with interpreter required, coded as [1] yes, [0] no.

Analytic strategy

An extended Cox proportional hazards model (conditional model) (Prentice, Williams, & Peterson, 1981) was fitted to the data. This model was appropriate since our event of interest (exiting the welfare system) is conditional on having received income support in the previous period. Further, this model allowed us to measure the association of separation and the timing of exiting the welfare system. We included further covariates to assess the impact of these characteristics on exiting the welfare system, such as age, gender, Indigenous status, refugee status, non-English speaking background, and receipt of DSP.

In this formulation, the logarithm of the hazard is modelled as a function of the baseline hazard and the selected covariates. The hazard represents the probability of an event (exit from income support) occurring among those individuals at time t who have not yet experienced the event (Allison, 2014).

The shape of the baseline hazard $a(t)$, is a function of duration in receipt of income support payments at time t when the values of all substantive explanatory variables are zero.

A hazard model with two explanatory variables, one constant and one time-variant, can be written as:

$$\log h(t) = a(t) + b_1x_1 + b_2x_2(t)$$

where $a(t)$ is any function of time, and the hazard at time t , $h(t)$ (i.e. the hazard of exiting the welfare system), is conditional on the values of x_1 and x_2 at the same time t .

In the Cox proportional hazards model the ratio of the hazards, c_k , for any two individuals at any point in time is constant and expressed as:

$$\frac{h_{ik}(t)}{h_{jk}(t)} = c_k$$

where i and j refer to two distinct individuals, and the constant c_k , does not depend on time, but may vary across explanatory variables k , such as gender, age and indigenous status (Allison, 2014).

Each person can receive income support during multiple non-consecutive periods of time and hence can contribute to more than one income support period in the pool of observations. To adjust for this ‘clustering’ of observations and to correct for statistical dependence robust standard errors were computed using the ‘sandwich’ method of Huber (1967) and White (1980).

Censoring

In event-history analysis, the term censoring is used to describe incomplete data for an individual during the observation period. While we had complete information on individuals’ income support history, due to the observation period ending on 30 June 2013, we had artificially cut the individuals time on income support so that the end of the income support spell did not correspond to an exit from the income support system. Hence, individuals in this dataset were (right) censored when they had not exited the welfare system by the 30 June 2013. Individuals who were deceased during the period of receiving income support were also (right) censored. The applied model accounts for right censoring by assuming that the censoring times are non-informative.

Results

Descriptive analysis

The population of participants in the analytic dataset includes $N = 537,912$ individuals and shows different characteristics on key demographic variables when compared to the total Australian population. Table 1 shows that 62.2% were females (compared to 50.6% in 2011 (Australian Bureau of Statistics, 2011b)), 66.9% were aged between 35 and 54 years (compared to 28.7% in 2011 (Australian Bureau of Statistics, 2011b)), 6.4% identified as Indigenous (compared to 3% in 2011 (Australian Bureau of Statistics, 2011c)), 3.4% were refugees and 12.4% required an interpreter (compared to 2.4% in 2011 speaking English not well or not at all (Australian Bureau of Statistics, 2011a)). During the observation period, 4% (22,438) of individuals were recorded as deceased. The 537,912 individuals included in the analysis were observed for 855,181 spells of income support receipt of which 184,683 (21.6%) corresponded to receipt of DSP.

Further, Table 1 shows the differences on the key demographic variables of the administrative dataset with HILDA. In comparison to the administrative dataset, the HILDA data has a higher percentage of female individuals, (HILDA: 67.2%; administrative dataset: 62.2%), the sample is overall younger (aged 14–34 years: HILDA: 43.7%; administrative dataset: 33.1%), there is a slightly lower percentage of individuals who identified as Indigenous (HILDA: 4.4%; administrative dataset: 6.4%); a higher percentage that were refugees (HILDA: 14.2%; administrative dataset: 3.4%); and a lower percentage of individuals that require an interpreter or have received DSP (interpreter required: HILDA: 2.8%; administrative dataset: 12.4%; DSP: HILDA: 14.2%; administrative dataset: 21.6%). In the HILDA data 2.3% deceased compared to 4.2% in the administrative dataset.

Event-history model

We constructed event-history models to analyse the administrative data and the HILDA data separately. There were considerable differences in the structure of the administrative data and the panel survey data and hence the model specifications and results are not comparable. Firstly, the

Table 1. Frequency of individuals by demographic characteristics for Administrative and HILDA data.

Variable	Categories	Administrative Dataset		HILDA	
		N	%	N	%
<i>Gender</i> ^{#H}	female	334,550	62.19	560	67.15
	male	203,362	37.81	274	32.85
<i>Age group</i> ^{#H}	14–24	48,540	9.02	127	15.23
	25–34	129,586	24.09	237	28.42
	35–44	170,878	31.77	287	34.41
	45–54	188,908	35.12	183	21.94
<i>Indigenous status</i> ^H	indigenous	34,141	6.35	37	4.44
	non-indigenous	503,771	93.65	616	73.86
	missing	n/a	n/a	181	21.7
<i>Refugee</i> ^{#* Z}	Yes	18,134	3.37	118	14.15
	No	519,778	96.63	17	2.04
	missing	n/a	n/a	699	83.81
<i>Interpreter required</i> ^{* H}	Yes	65,535	12.43	23	2.76
	No	461,486	87.57	804	96.4
	missing	n/a	n/a	7	0.84
<i>received DSP</i> ^{## H}	Yes	184,683	21.6	118	14.15
	No	670,498	78.4	414	49.64
	missing	n/a	n/a	302	36.21
<i>Deceased</i>	Yes	22,438	4.17	19	2.28
	No	515,474	95.83	815	97.72

[#] N = 10,891 individuals not asked;

^{##} not per individual but defined as per spell (total number of spells = 855,181);

* as of 30 June 2003.

information in the administrative data on income support recipients is updated fortnightly in the system, whereas the panel survey data is collected by self-report annually. This means that the start and endpoint of receiving income support is not recorded with as much detail. Secondly, while subgroups of interest were well represented in the administrative data, the panel survey data did not include sufficient numbers of these subgroups in the sample and model effects could not be estimated (see Supplemental Material – HILDA data analysis for detail). The results from the event-history model applied to the administrative data are described below.

Table 2 shows the results from the event-history model fitted to the administrative data on exit from income support. Four models were fitted to the data (Models 1–4) with each model building on the previous model by including additional explanatory variables. All models included the baseline hazard function but only the estimated hazard ratios (HR) for the explanatory variables on exit from income support are shown in Table 1. Model 1 is the simplest model and includes only relationship status (non-partnered versus partnered) as the explanatory variable. The hazard ratio estimated for Model 1 is the odds of exiting income support in any given quarter for individuals who are no longer partnered relative to those who have remained in a partnership since commencing receipt of income support, without controlling for any other demographic variables. Model 2 builds on Model 1 by including gender and age as explanatory variables in addition to relationship status. Model 3 builds on Model 2 by including Indigenous status, whether an interpreter is required or not and whether the individual identifies as a refugee. The final Model 4 includes an indicator identifying whether DSP was received during the spell. The first column in Table 1 includes the variable name and the second column includes the categories of the variables. The third column shows the estimated hazard ratio for Model 1, followed by the 95% hazard ratio confidence limits. The remaining columns show the estimated hazard ratios and corresponding 95% confidence intervals for the variables in Models 2–4.

Without controlling for any other demographic variables, results for Model 1 show that individuals who separate from their partner are marginally less likely to exit income support (HR = 0.97) compared to those that remain partnered. While this ratio is very close to one, the 95% confidence limits indicate that this finding is statistically significant at the 5% level. With

gender and age added as explanatory variables, the results for Model 2 show that individuals who are no longer partnered (HR = 0.86), and females (HR = 0.70), are less likely to exit income support. Younger individuals aged 14–24, 25–34 and 35–44 are more likely to exit income support than those aged 45–54 years old (hazard ratios >2.0). The estimates for Model 3 additionally show that individuals of Indigenous status (HR = 0.82), those who required an interpreter (HR = $1/1.46 = 0.68$) and identify as a refugee (HR = $1/1.14 = 0.88$), are also less likely to exit income support.

In the final Model 4, an indicator for DSP is included to examine the change in the hazard ratio for relationship status after controlling for this typically long-term payment. As expected, the hazard ratio for individuals in receipt of DSP is high (HR = 7.74) and the hazard ratio for individuals who are no longer partnered remains stable at HR = 0.87. The hazard ratio for females to exit income support has decreased to HR = 0.54 (from 0.69 in Model 3). Indigenous status is also associated with a lower hazard of exiting income support compared to non-Indigenous individuals (HR = 0.78) and younger individuals aged 14–24, 25–34 and 35–44 are more likely to exit than those aged 45–54 years old; however, the hazard ratios have decreased from those estimated in Model 3 to 1.72, 1.71 and 1.65, respectively.

Table 1 has shown that while the overall estimated hazard of exiting income support is marginally but significantly lower for individuals who become separated following entry to income support (HR = 0.97), this association becomes more pronounced when gender and age are also considered (HR = 0.86). While each of the four variables representing Indigenous status, the requirement for an interpreter, refugee status and DSP indicator are also associated with the hazard of exiting income support the effect of relationship status remains stable when they are added to the model following gender and age.

Discussion

This project was a collaboration between the DHS and the Life Course Centre with three main objectives: First to trial a model of accessing and analysing a full population administrative dataset. Second to illustrate the use of administrative data for research by examining a substantive question on the impact of relationship breakdown on the duration of income support receipt. And finally, to further assess the potential of administrative data for research. Overall, we explored a process by which a team of university-based researchers could effectively partner with an Australian Government agency to securely generate research evidence from administrative data that records income support payment information for the entire population in receipt of this benefit.

In order to securely access and analyse the dataset, one member of the academic team worked from the data custodian's premises. This enabled the data analysis, but also facilitated the data custodian as an active collaborative partner. Working in close collaboration with the data custodian was vital particularly during the data management period. We found that university researchers and data custodian's personnel apply different work strategies and routines and both parties adjusted to facilitate a successful project outcome. Any outputs produced by the university researcher required clearance by several personnel before it was permitted to be taken out of the data custodian's premises and shared with the remaining university team for discussion. Open and frequent continuing communication was key to ensuring that an accurate statistical model was applied to the data. This guaranteed that realistic results were obtained and interpreted correctly. The collaboration with the data custodian was central to this project, given their thorough knowledge of the welfare payment system and eligibility requirements for different income support payment types, their familiarity with the structure and contents of the dataset, data collection and storage processes. It is critical that social scientists seeking to analyse administrative data for research undertake a collaborative approach with data custodians in the identification and development of research projects to produce policy-relevant outputs. It is not the function of a government department to create and manage datasets designed for a range of potential research questions

and therefore other researchers interested in accessing similar data need to follow similar steps and establish new partnerships. It would be expected that the more experience government personnel and researchers gain in working together and documenting the approach, the more streamlined these steps for accessing and analysis of administrative data for addressing important policy questions may become.

We demonstrated the use of administrative data for social research by applying statistical methods for longitudinal data to a case study examining the association between relationship breakdown and duration of income support receipt. We found that relationship breakdown, following initial receipt of income support payments while in a partnered relationship, was associated with a significantly lower likelihood of exiting payment receipt. This result held while controlling for other factors known to have an impact on the use of income support including gender, age, Indigenous status, whether an interpreter is required, refugee status and whether or not the individual has received a disability pension. The Australian Government Department of Social Services has been investing in research to further understand the circumstances of individuals in receipt of income support in order to develop evidence-based policies to improve employment opportunities (Department of Social Services, 2017). Through this process the Department identified groups of individuals who have particularly high lifetime costs that can be reduced through improved policy settings and targeted interventions. Our results showed that individuals in receipt of income support are less likely to exit the welfare system following a relationship breakdown compared to their partnered counterparts. Furthermore, we found that females are almost half as likely to exit the welfare system relative to males. These results highlight the importance of policy for providing opportunities to enter employment following separation for those who need it most, including women, refugees, individuals who require an interpreter or identify as Aboriginal or Torres Strait Islander.

The study has highlighted the importance of longitudinal administrative data for social science research and some of the strengths and weaknesses of this type of data compared to national panel survey data. The administrative data includes the full population of income support recipients and hence is particularly useful when the research is focused on minority groups and hard-to-reach populations. For example, if we were to select people receiving income support from a national panel survey the number of observations would be too small to allow investigation of differences, particularly if further disaggregated by other characteristics such as Indigenous and refugee status. We attempted to conduct the identical analysis using a nationally representative household panel survey; however, as the panel survey data was designed to provide a representation of households from the national population rather than of households in receipt of a government service, we were not able to reproduce the analysis due to underrepresentation of minority groups in our cohort of interest.

On the other hand, due to the size of the administrative dataset, data management was extensive and the processing times for the statistical models were lengthy. As the data is not collected for research, some of the key demographic variables were not recorded in a consistent way as we would expect from well-managed longitudinal surveys and some potential covariates are not included at all. This is a limitation to assess the sensitivity and robustness of findings from statistical models. Marital status, for example, is often measured in surveys to distinguish between married, cohabiting, divorced, separated, widowed and single states. However, this level of detail is not required for records in the administrative context so detailed marital transitions cannot be accounted for in analyses of these data. Another common feature of administrative data is that information is collected about the individual throughout the period of receipt of a service. When the service is no longer required, there is no further information available about this person which makes it difficult to investigate triggers that lead to service requirements and subsequent outcomes after service provision has been relinquished.

The absence of a wide range of covariates and the physical restrictions to accessing the data reduced the possibility of undertaking sensitivity analyses and assessing the robustness of

statistical models, typically undertaken to strengthen the findings. Further limitations included the considerable amount of data management in preparation for analysis, the unknown accuracy of the administrative data on variables that are not necessary for providing the service and the inability to reproduce the analysis of the data by others due to restricted data access. These challenges have also been reported by Connelly and Gayle (2017); Foley, Shuttleworth, and Martin (2018); Peabody, Luck, Jain, Bertenthal, and Glassman (2004); Playford, Gayle, Connelly, and Gray (2016).

Overall, our research highlights some of the strengths and limitations of administrative data compared to survey data and the importance of close collaboration with data custodians when analysing administrative data. While accessing sensitive data from within the premises of the data custodians was the preferred option of the data custodian, this posed several difficulties for the research team, including the limitation to use a specific statistical analysis software program, time-consuming travel to and from the data custodian's facilities and difficulties to meet and discuss output among the research team due to strict regulations regarding removing research outputs from the data custodian's office. While data security is of the highest priority, other access possibilities could be explored to overcome these hurdles without compromising confidentiality. Current directions and trends toward more open data access in Australia, as well as many other countries, suggest that new opportunities for realising the value of administrative data for research, as well as policy design and evaluation are becoming increasingly available. This is unlikely to negate the importance of continuing to collect rich longitudinal data from national survey samples, but it does open new possibilities for important new research and policy insights into hard-to-reach minority groups. For researchers concerned with understanding pathways into and out of disadvantage for such groups, access to administrative data is thus imperative and an exciting development in social science infrastructure.

Acknowledgments

This research was conducted with data from, and the support of, the Australian Government Department of Human Services (DHS). The views expressed in this paper are those of the authors alone and do not represent those of DHS. The authors would like to thank personnel from the Research and Analysis Section of DHS who have contributed.

This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to either DSS or the Melbourne Institute.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the Australian Research Council (ARC) Centre of Excellence for Children and Families over the Life Course (ARC Grant ID: CE140100027).

Notes on contributors

Dr Melanie Spallek is a Research Fellow at the Institute for Learning Sciences and Teacher Education. Her research is centrally concerned with longitudinal and big administrative data to investigate life course trajectories, in particular school-to-work trajectories and related performance outcomes. Of special interest are certain disadvantaged sub-groups, such as welfare recipients, adolescents from out-of-home-care or with ASD. Her research applies innovative statistical methodology to explain trajectories and drivers of change for certain transitions. Through her involvement

with the Life Course Centre, she has published research on housing trajectories, home ownership transitions and family events over the life course using longitudinal data.

Michele Haynes is Professor of data analytics for education and social research at the Institute for Learning Sciences and Teacher Education. Michele is an experienced statistician with expertise in estimation of complex models for social applications using data from multiple sources including national panel surveys and government administrative data. Her research focus is on data innovation for longitudinal social research and improving statistical techniques for analysing pathways of individuals and families through disadvantage and identifying the drivers of change using longitudinal data. More recently, this includes analysing education performance and well-being outcomes for children through the school years and identifying the factors that trigger improvements and maintain performance. Previous research has investigated the effects of family dynamics on both housing transitions and welfare dependency.

Janeen Baxter is Professor of Sociology and Director of the Australian Research Council Centre of Excellence for Children and Families over the Life Course in the Institute for Social Science Research at The University of Queensland. Janeen conducts research in the areas of gender, families, households, social disadvantage and the life course. She is an elected Fellow of the Academy of Social Sciences in Australia and a former member of its Executive Committee. Janeen is currently leading the Evaluation of the Try, Test and Learn Fund for the Australian Government Department of Social Services and is a member of several government advisory committees and expert reference groups.

Nicole Kapelle is currently enrolled as a PhD candidate at the Institute for Social Science Research at The University of Queensland, Australia, and holds a position as Research Associate at the Department of Social Sciences at Humboldt-Universität zu Berlin, Germany. For her PhD, Nicole explores the economic consequences of partnership dissolutions. More precisely, she is interested in how personal wealth accumulation of men and women is disrupted through the dissolution of a partnership. Within her position at Humboldt-Universität zu Berlin, she explores the emergence and development of within-couple wealth inequalities over time in marriage. For her research, Nicole uses longitudinal data and applies methods such as sequence analysis, latent class analysis, and fixed effects and random effects regression models.

ORCID

Melanie Spallek  <http://orcid.org/0000-0001-8364-4896>

Michele Haynes  <http://orcid.org/0000-0003-2718-3954>

Janeen Baxter  <http://orcid.org/0000-0002-8723-9000>

Nicole Kapelle  <http://orcid.org/0000-0001-5855-1153>

References

- Allison, P. (2014). *Event history and survival analysis*. Los Angeles, USA: Sage Publications.
- Andreß, H.-J., Borgloh, B., Bröckel, M., Giesselmann, M., & Hummelsheim, D. (2006). The economic consequences of partnership dissolution—A comparative analysis of panel studies from Belgium, Germany, Great Britain, Italy, and Sweden. *European Sociological Review*, 22(5), 533–560.
- Australian Bureau of Statistics. (2011a). 2011 census community profiles - Basic community profile. Retrieved from http://www.censusdata.abs.gov.au/census_services/getproduct/census/2011/communityprofile/0?opendocument&navpos=220
- Australian Bureau of Statistics. (2011b). 2011 Census Quick Stats. Retrieved from http://www.censusdata.abs.gov.au/census_services/getproduct/census/2011/quickstat/0?opendocument.
- Australian Bureau of Statistics. (2011c). Estimates of Aboriginal and Torres Strait Islander Australians, June 2011 (Cat. no. 3238.0.55.001). Retrieved from <https://www.abs.gov.au/ausstats/abs@.nsf/lookup/3238.0.55.001Media%20Release1June%202011>
- Bane, M. J., & Ellwood, D. T. (1983). *The dynamics of dependence: The routes to self-sufficiency*. Cambridge, Massachusetts: Urban Systems Research and Engineering, Incorporated.
- Barrett, G. F. (2000). The effect of educational attainment on welfare dependence: Evidence from Canada. *Journal of Public Economics*, 77(2), 209–232.
- Bayaz-Ozturk, G., Burkhauser, R. V., Couch, K. A., & Hauser, R. (2018). The effects of union dissolution on the economic resources of men and women: A comparative analysis of Germany and the United States, 1985–2013. *The ANNALS of the American Academy of Political and Social Science*, 680(1), 235–258.
- Blank, R. M. (1989). Analyzing the length of welfare spells. *Journal of Public Economics*, 39(3), 245–273.

- Blank, R. M., & Ruggles, P. (1994). Short-term recidivism among public-assistance recipients. *The American Economic Review*, 84(2), 49–53.
- Bora, C. E., Caudill, P. J., Spera, C., & Kunz, J. F. (1998). A look at life after welfare. *Public Welfare*, 56, 32–37.
- Cheng, T. (2002). Welfare recipients: How do they become independent? *Social Work Research*, 26(3), 159–170.
- Cipollone, A., Patacchini, E., & Vallanti, G. (2014). Female labour market participation in Europe: Novel evidence on trends and shaping factors. *IZA Journal of European Labor Studies*, 3(1), 18.
- Coe, R. D. (1981). A preliminary empirical examination of the dynamics of welfare use. *Five Thousand American Families: Patterns of Economic Progress*, 9, 121–168.
- Connelly, R., & Gayle, V. (2017). An investigation of the consistency of parental occupational information in UK Birth Records and a National Social Survey. *European Sociological Review*, 33(2), 240–256.
- Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, 1–12.
- Craig, L., Mullan, K., & Blaxland, M. (2010). Parenthood, policy and work-family time in Australia 1992–2006. *Work, Employment and Society*, 24(1), 27–45.
- Crichton, S., Templeton, R., & Tumen, S. (2015). *Using integrated administrative data to understand children at risk of poor outcomes as young adults*. Wellington New Zealand: New Zealand Treasury.
- De Vaus, D., Gray, M., Qu, L., & Stanton, D. (2014). The economic consequences of divorce in Australia. *International Journal of Law, Policy and the Family*, 28(1), 26–47.
- De Vaus, D., Gray, M., Qu, L., & Stanton, D. (2015). *The economic consequences of divorce in six OECD countries*. Melbourne. Retrieved from <https://aifs.gov.au/publications/economic-consequences-divorce-six-oecd-countries>
- Department of Social Services. (2017). *30 June 2016 Valuation Report*. Canberra, Australia.
- Department of Social Services, Department of Agriculture and Water Resources, Department of Employment, & Department of Education and Training. (2017). *A guide to Australian Government payments*. Canberra, Australia.
- Elias, P. (2014). *Administrative data. Facing the Future: European Research Infrastructures for the Humanities and Social Sciences*. Berlin, Germany.
- Foley, B., Shuttleworth, I., & Martin, D. (2018). Administrative data quality: Investigating record-level address accuracy in the Northern Ireland Health Register. *Journal of Official Statistics*, 34(1), 55–81.
- Gaudet, S., Cooke, M., & Jacob, J. (2011). Working after childbirth: A lifecourse transition analysis of Canadian women from the 1970s to the 2000s. *Canadian Review of Sociology/Revue Canadienne De Sociologie*, 48(2), 153–180.
- Gault, B., Hartmann, H., & Yi, H.-Y. (1998). Prospects for low-income mothers' economic survival under welfare reform. *Publius: the Journal of Federalism*, 28(3), 175–193.
- Harris, K. M. (1993). Work and welfare among single mothers in poverty. *American Journal of Sociology*, 99(2), 317–352.
- Harron, K., Dibben, C., Boyd, J., Hjern, A., Azimae, M., Barreto, M. L., & Goldstein, H. (2017). Challenges in administrative data linkage for research. *Big Data & Society*, 4(2), 2053951717745678.
- Huber, P. J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions*. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. University of California, Berkeley, California.
- Kroch, E. A., & Sjoblom, K. (1994). Schooling as human capital or a signal: Some evidence. *Journal of Human Resources*, 29, 156–180.
- Leahy, P. J., Buss, T. F., & Quane, J. M. (1995). Time on welfare. *American Journal of Economics and Sociology*, 54(1), 33–46.
- Maroto, M. L. (2015). Pathways into bankruptcy: Accumulating disadvantage and the consequences of adverse life events. *Sociological Inquiry*, 85(2), 183–216.
- McKeever, M., & Wolfinger, N. H. (2001). Reexamining the economic costs of marital disruption for women. *Social Science Quarterly*, 82(1), 202–217.
- Meyer, D. R., & Cancian, M. (1998). Economic well-being following an exit from Aid to Families with Dependent Children. *Journal of Marriage and the Family*, 60, 479–492.
- Meyers, M. K., & Heintze, T. (1999). The performance of the child-care subsidy system. *Social Service Review*, 73(1), 37–64.
- Pavetti, L. A. (1994). *The dynamics of welfare and work: Exploring the process by which women work their way off welfare*. Malcolm Wiener Center for Social Policy, Working Papers: Dissertation Series (#D-93-1), Harvard University, Cambridge, Massachusetts.
- Peabody, J. W., Luck, J., Jain, S., Bertenthal, D., & Glassman, P. (2004). Assessing the accuracy of administrative data in health information systems. *Medical Care*, 42(11), 1066–1072.
- Pearce, D. (2000). Rights and wrongs of welfare reform: A feminist approach. *Affilia*, 15(2), 133–152.
- Petersen, C. D. (1995). Female-headed families on AFDC: Who leaves welfare quickly and who doesn't. *Journal of Economic Issues*, 29(2), 619–628.
- Piskulich, C. M. (1993). *Toward a comprehensive model of welfare exits: Aid to Families with Dependent Children, Food Stamps and Medicaid*. *American Journal of Political Science*, 37(1), 165–185.

- Playford, C. J., Gayle, V., Connelly, R., & Gray, A. J. (2016). Administrative social science data: The challenge of reproducible research. *Big Data & Society*, 3(2), 2053951716684143. Retrieved from: <https://journals.sagepub.com/doi/abs/10.1177/2053951716684143>
- Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2), 373–379.
- Productivity Commission. (2017). *Data Availability and Use: Overview & Recommendations*. Canberra. Retrieved from <https://www.pc.gov.au/inquiries/completed/data-access/report/data-access-overview.pdf>
- SAS Institute Inc. (2013). *SAS and all other SAS Institute Inc. product and service names are registered trademarks or trademarks of SAS Institute Inc.* Cary, NC, USA
- Sheehan, G. (2002). Financial aspects of the divorce transition in Australia: Recent empirical findings. *International Journal of Law, Policy and the Family*, 16(1), 95–126.
- Smyth, B., Rodgers, B., Son, V., & Vnuk, M. (2015). The Australian child support reforms: A critical evaluation. *Australian Journal of Social Issues*, 50(3), 217–232.
- Smyth, B., & Weston, R. (2000). *Financial living standards after divorce: A recent snapshot* Melbourne. Retrieved from <https://aifs.gov.au/publications/financial-living-standards-after-divorce>
- Stellmack, A. L., Wanberg, C. R., & Kammeyer-mueller, J. (2003). Transitions off welfare: An examination of demographic, socioeconomic, and motivational predictors. *Industrial Relations: A Journal of Economy and Society*, 42(4), 623–649.
- Tach, L. M., & Eads, A. (2015). Trends in the economic consequences of marital and cohabitation dissolution in the United States. *Demography*, 52(2), 401–432.
- Tamborini, C. R., Couch, K. A., & Reznik, G. L. (2015). Long-term impact of divorce on women's earnings across multiple divorce windows: A life course perspective. *Advances in Life Course Research*, 26, 44–59.
- Tseng, Y.-P., Vu, H., & Wilkins, R. (2008). Dynamic properties of income support receipt in Australia. *Australian Economic Review*, 41(1), 32–55.
- Tseng, Y.-P., & Wilkins, R. (2003). Reliance on income support in Australia: Prevalence and persistence. *Economic Record*, 79(245), 196–217.
- Van Damme, M. C. D. (2010). *Beyond marriage: Women's economic independence and separation in comparative perspective* PhD thesis. Tilburg University, Tilburg.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838.
- Wilson, M. (1999). The duration of benefit receipt: New findings from the benefit dynamics data set. *Social Policy Journal of New Zealand*, 59(13).