The Road Less Travelled: Using Administrative Data to Understand Inequalities by Sexual Orientation

Francisco Perales iD , Associate Professor of Sociology at the School of Social Science

The University of Queensland, Queensland, Australia

ABSTRACT

Understanding the processes contributing to equality of opportunity and outcomes in contemporary societies is at the core of the discipline of sociology. This paper illustrates the value of administrative data to underpin research aimed at identifying, monitoring, and addressing socio-economic disparities between population groups. To accomplish this, I draw on three case studies of recent empirical research leveraging administrative data to examine processes contributing to the (re)production of inequalities by sexual orientation. Collectively, the three case studies exemplify how data sources that fall within the broad category of 'administrative data' can help social researchers generate new, policy-relevant knowledge on socio-economic inequalities, as well as robust information to contextualize public and legislative debate. The paper concludes with a discussion of the promises and challenges of using administrative data to understand inequalities by sexual orientation, as well as inequalities between other minority and majority groups.

Keywords –Equality, transparency, positive duties, mainstreaming

Summary

1. INTRODUCTION

Understanding the processes underpinning equality of opportunity and outcomes in contemporary societies is at the core of the social sciences in general, and the discipline of sociology more specifically (Mullan 2017). Identifying the structural drivers of socio-economic inequalities is an important step towards redressing them through the implementation of appropriate evidence-based policies, practices and interventions. This mode of thinking about inequality fits squarely within sociologists' renewed interest in solution-oriented research developed for and with stakeholders—including local and national government, philanthropic think-tanks, and not-for-profit organizations (Watts 2017; Western 2019).

Improving our knowledge of the factors that contribute to establishing, entrenching and dissipating socio-economic inequalities is nevertheless contingent on the availability of fit-for-purpose datasets that can be interrogated to generate accurate and reliable answers. Traditional data sources used in sociological research—and social science research, more broadly—involve both qualitative and quantitative instruments. Qualitative approaches include in-depth interviews, focus groups and ethnographic observation, whereas quantitative approaches often rely on experiments and social surveys (Bryman 2016). This paper is concerned with quantitative forms of inquiry. Within this paradigm, a recent movement involves the utilization of administrative data sources to understand and address processes of socio-economic inequality (Chetty et al. 2017; Einav and Levin 2014; Picketty and Saez 2014). Here, I adopt a broad definition of administrative data, as information derived from the operation of administrative systems for purposes other than research (Connelly et al. 2016).

The aim of this contribution is to illustrate the value of administrative datasets to underpin research aimed at understanding and addressing socio-economic inequalities (Australian Government Productivity Commission 2013). To accomplish this, I will rely on three case studies of recent research that I have undertaken together with colleagues and which focused on examining socio-economic disadvantage amongst individuals who identify as a sexual minority (e.g., as gay, lesbian, bisexual, or queer).[1] These case studies encompass empirical research on the processes contributing to the production and reproduction of inequalities by sexual orientation using administrative datasets, with findings that hold potential to inform policies aimed at redressing these inequities.

The remainder of this article is structured as follows. In the next section, I briefly summarize social science knowledge on socio-economic inequalities by sexual orientation. Subsequently, I will discuss the potential of administrative data as sources of information to study processes of inequality in general, and inequalities by sexual orientation more specifically. A section introducing the three case studies follows. The paper concludes with a discussion of the promises and perils of using administrative data to understand inequalities by sexual orientation, as well as inequalities between other minority and majority groups.

2. EQUALITY OF OPPORTUNITY AND OUTCOMES BY SEXUAL ORIENTATION

A long-standing body of evidence within the social sciences has been devoted to measuring and making sense of socio-economic inequalities structured along the axes

of gender, race/ethnicity, migrant status and social class. Fewer contributions have addressed Inequality by sexual orientation, with the paucity of such research being often attributed to the absence of suitable data (see e.g., Waite and Denier 2019). As such, social scientists have only recently begun to systematically document how individuals' sexuality—captured by their sexual identity, attractions and/or behavior—influences their outcomes across different domains of life. Existing empirical studies have largely focused on health outcomes, a legacy of the classification of homosexuality as a mental illness up until the mid-1970s (Drescher 2015) and of the AIDS epidemic being the focus of sexual-orientation research over the 1980s and 1990s (Institute of Medicine 2011). The available evidence demonstrates that, across the globe, individuals who identify as a sexual minority tend to experience poorer physical and mental health than individuals who identify as heterosexual, *ceteris paribus* (Institute of Medicine 2011; Perales 2016, 2019; Perales and Campbell 2019).

The dominant paradigm used to explain these disparities is the minority stress model, which posits that structural stigma towards sexual minorities constitutes the main factor contributing to their suboptimal health outcomes (Meyer 2003, Hatzenbuehler 2009). Traditional, hetero-normative societal scripts portray non-heterosexuality as an undesirable, subordinate, and/or pathological status, with a sizeable share of the population in developed and developing countries alike still endorsing such views (Perales and Campbell 2018; Valfort 2017). Because of this, individuals whose identities, attractions and/or behaviors fall outside the heterosexual norm are subjected to unique stressors over the course of their day-to-day lives. Such stressors can be proximal (e.g., internalized homophobia, identity concealment, and rejection expectation) or distal (ranging from subtle micro-aggressions to overt instances of interpersonal violence).

Stressors are argued to get 'under the skin' by generating unhealthy psycho-biological stress responses in the body, inhibiting help-seeking, and fostering maladaptive coping behaviors (e.g., social withdrawal, substance abuse, and self-harm) amongst individuals from

sexual minorities (Meyer 2003, Hatzenbuehler 2009, Pachankis et al. 2015). The minority stress model therefore stands in direct opposition to 'deficit' approaches to explaining LGBTIQ+ disadvantage. Such approaches—more often found in political rhetoric and popular culture than in academic scholarship—attribute health disparities favoring the heterosexual majority to the lifestyle choices or psychological traits of sexual-minority people.

While the minority stress model was developed to conceptualize sexual-orientation health disparities, it has subsequently been refined and re-deployed to explain socio-economic inequalities by sexual orientation in other life domains. This includes evidence that, relative to their heterosexual counterparts, individuals from sexual minorities are more likely to encounter income poverty and material deprivation, face spells of homelessness, be subjected to workplace exclusion and barriers to career progression, and experience family rejection and estrangement (Badgett et al. 2019; Fraser et al. 2019; Perales et al. 2020; Uhrig 2015). These inequalities begin early into the life course (Perales and Campbell 2019), vary across communities, regions, and countries with differing levels of sexual stigma (Pachankis et al. 2015), and are more pronounced for certain sexual-minority groups—such as bisexual people—than others (Perales 2019).

Given its depth and breadth, the disadvantage experienced by individuals from sexual minorities is an important object of academic study. The empirical evidence that social scientists generate can play an important role in the development, testing and refinement of policies and programs aimed at enhancing equality of opportunity and outcomes by sexual orientation (Institute of Medicine 2011). Further, social science scholarship is sometimes sought and used in legislative processes concerning the rights of sexual minorities (Gates 2015; Manning, Fettro and Lamidi 2014). Despite the sore need for robust evidence in this space, the absence of suitable data often restricts the scope and quality of empirical studies on the intersections between inequality and sexual orientation (Institute of Medicine 2011; Waite and Denier 2019). In this context, administrative

datasets stand as a promising—yet infrequently used—alternative to traditional social surveys.

3. THE PROMISE OF ADMINISTRATIVE DATA FOR INEQUALITY RESEARCH

As part of the 'big data' revolution, there is growing recognition of the value of administrative data for social research (Card et al. 2010; Connelly et al. 2016). An increasingly popular type of administrative data for inequality research are official records on individuals, their families and their households collected by government and made available to researchers for analysis through safe and controlled processes. This includes, for example, data from a country's population Census, social security data containing income-support records, datasets collating information from schools or higher-education providers, and data capturing individuals' interactions with the health-care system. Administrative datasets of interest to scholars of inequality may also come in other forms, such as records on service delivery collected by non-for-profit organizations and other frontline agencies (Ambrey et al. 2019; Parsell et al. 2017) or aggregate-level information stemming from elections, ballots, plebiscites, and referenda (Gravelle and Carlson 2019; Perales and Todd 2018). Linkages of multiple administrative data sources (e.g., register data on demographic processes and income tax data) further enrich the information available for each individual record and multiply the analytic pathways.

Administrative data can offer multiple and unique analytic benefits to inequality researchers, which can significantly lift the scope, quality and impact of their research outputs. First, administrative datasets are generally much larger in volume than traditional social surveys; sometimes, they contain information on complete populations (Card et al. 2010). Within the social sciences, they may fall in the category of 'big data' (Connelly et al. 2016). As a result, certain administrative data permit detailed and accurate analyses of small populations of key policy interest—such as refugees, ethnic minorities, or same-sex couples. Because of their large

numbers, administrative data are also fit-for-purpose to follow intersectional perspectives (McCall 2005) and consider the compounding effects on individual outcomes of detailed permutations of disadvantaged social statuses.

Second, administrative datasets are frequently collected over a prolonged period of time (Card et al. 2010). This enables researchers to track individuals' circumstances and outcomes over the long term, allowing for more nuanced understandings of how inequality accumulates and compounds across generations and over the course of life (Chetty et al. 2017; DiPrete and Eirich 2005; Picketty and Saez 2014). Besides, such data is often available retrospectively. This is a significant advantage over longitudinal survey datasets, which require a lengthy maturation period—with the most insightful analyses having to wait until a sufficient amount of information has been collected from the sample (Australian Government Productivity Commission 2013). Large and long-running administrative data collections also open opportunities for researchers to assess historical or generational change in inequality-generating processes, to leverage changes in macro-level conditions as natural experiments to tease out causal relationships, or to undertake long-term follow-ups of natural experiments (Card et al. 2010, Connelly et al. 2016, Australian Government Productivity Commission 2013).

Third, administrative datasets usually contain official, verifiable records. This results in smaller reporting errors and biases than those arising when collecting objective information using social surveys (Card et al. 2010; Australian Government Productivity Commission 2013). Take for example measures on annual income. Standard social surveys rely on individuals' accurately recollecting and reporting complex information on different income sources and undertaking complex mental calculations (e.g., adding amounts or deducting taxes). Sometimes, survey respondents are required to estimate the income of other household members, which incurs an even greater risk of error. Administrative datasets, on the other hand, may simply obtain annual-in

come information from employer databases or tax records at the individual or household level. The same can be said of information on, for example, individuals' specific interactions with the income-support or healthcare system, long-term residential and migration histories, or detailed educational performance and attainment over the life course (for concrete examples, see Short et al. 2009 and Drapeau et al.2011). Constructs such as these are of great interest to inequality scholars.

Finally, administrative data is relatively inexpensive, as there are no extra costs associated with their collection—although costs towards preparing the data for analysis and establishing a secure data-access system need to be factored in. Since administrative data often represent 'value for money', it can be argued that their use for research purposes constitutes an ethical alternative to allocating hefty amounts of tax-payer money to the collection of large volumes of primary data. From this perspective, administrative datasets can be seen as a public good (Mervis 2017).

Studies examining overall processes of inequality such as income mobility and economic inequality (Chetty et al. 2017; Einav and Levin 2014; Picketty and Saez 2014)—are beginning to harness the potential of linked administrative datasets to advance the field in areas that had stalled due to data shortcomings. In the same vein, administrative data are enviably placed to study differences in opportunity structures for majority and minority groups, including differences between sexual minorities and the heterosexual majority. However, with some exceptions (see e.g., Andersson et al. 2006; Wiik, Seierstad and Noack 2014), few studies to date have utilized administrative data sources to improve our understanding of processes of socio-economic inequality by sexual orientation. In the next section, I outline three case studies that demonstrate the untapped potential of these data to expand the evidence base.

4. CASE STUDIES

4.1 CASE STUDY 1: USING AUSTRALIAN ADMINISTRATIVE DATA TO UNDERSTAND THE ROLE OF STIGMA AS A DRIVER OF SEXUAL-ORIENTATION HEALTH DISPARITIES

In late 2017, the Australian Government commissioned the Australian Bureau of Statistics (ABS) to conduct a national vote on same-sex marriage. This so-called *Australian Marriage Law Postal Survey* (AMLPS) was devised to inform potential legislative change to allow same-sex couples to marry. All Australian citizens who were on the Commonwealth Electoral Roll and at least 18 years of age were invited to participate by answering the question: *"Should the law be changed to allow same-sex couples to marry?"*. Approximately 79.5% of eligible Australians (~12.7 million people) participated in the survey, with a majority of 'Yes' votes (~61.6%, ~7.82 million) over 'No' votes (~38.4%, ~4.87 million). The AMLPS was criticized on multiple grounds—for example, for subjecting the LGBTIQ+ community to unnecessary stress and social scrutiny and for its high implementation costs (~AU$80.5 million). Despite this, an indirect benefit of the AMLPS was that it left behind a rich trail of administrative data with immense potential to advance research on social inequalities by sexual orientation.

The ABS did not make any data on individual votes in the AMLPS available for research purposes. However, it released a wealth of information on the aggregated vote results across geographical locations, including Australia's 150 Electoral Divisions. In a collaborative study, my colleague and I leveraged this aggregate-level information to approximate the notion of structural stigma (Perales and Todd 2018). Specifically, we exploited the fact that the percentage of 'No' votes out of all registered votes ranged markedly across Electoral Divisions, from 16.3% (City of Melbourne) to 73.9% (Blaxland, New South Wales). Areas with *higher* shares

of 'No' votes were identified as having higher levels of structural stigma, whereas areas with *lower* shares of 'No' votes were identified as having lower levels of structural stigma.

This aggregate-level measure of structural stigma was then linked to 2016 individual-level survey data from the *Household, Income and Labour Dynamics in Australia (HILDA) Survey*, on the basis of the Electoral Division in which each of the survey respondents lived. The HILDA Survey is a national probability survey of Australian households, and collects detailed information on respondents' socio-demographic circumstances and outcomes—including measures of their sexual orientation and health/wellbeing. Using the HILDA Survey data augmented with the AMLPS information, we were able to test a simple yet powerful premise: whether the health/wellbeing of individuals who identified as lesbian gay, or bisexual (LGB) varied depending on their residence in low-stigma compared to high-stigma areas.

Our results were highly illuminating and evidenced that structural stigma—and not LGB individuals themselves—is to blame for the poorer health and wellbeing observed amongst Australian LGB populations. LGB people living in areas with very low levels of stigma (i.e., with very few 'No' voters) exhibited health/wellbeing outcomes that were virtually undistinguishable from those of their heterosexual neighbors. In contrast, LGB people in areas with very high levels of stigma (i.e., with large shares of 'No' voters) exhibited much poorer outcomes than their heterosexual peers. A visual representation of the study results is presented in Figure 1. These findings are policy-relevant, suggesting that structural-level interventions to reduce stigma towards LGB populations are needed to close gaps in health/ wellbeing outcomes (Perales and Todd 2018).

This case study highlights how the decision by the ABS to release aggregate-level administrative data on the AMLPS results enabled researchers to make important inroads towards understanding inequalities by sexual orientation. In this instance, the new information was unique in its nature and difficult to replicate through traditional primary data-collection methods.

This collection of data was subsequently deployed in other research studies to further advance research on Australian sexual minorities (see Gravelle and Carlson 2019; Saxby et al. 2020; Wilson et al. 2020).
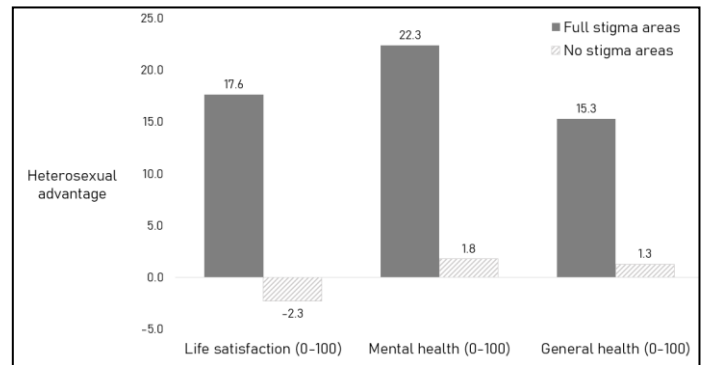
FIGURE 1: *Predicted health outcomes of lesbian, gay and bisexual people in 'full stigma' compared to 'no stigma' areas within Australia*

Notes: Based on analyses in Perales & Todd (2018). 'Full stigma areas' refers to hypothetical electoral divisions with 100% 'No' voters. 'No stigma areas' refers to hypothetical electoral divisions with 0% 'No' voters. All health/wellbeing outcomes are measured on a scale from 0 (worst health/wellbeing) to 100 (best health/wellbeing). The heterosexual advantage is calculated as the predicted health/wellbeing of an average heterosexual person minus the predicted health/wellbeing of an average lesbian, gay or bisexual person.

4.2 CASE STUDY 2: LEVERAGING DUTCH ADMINISTRA-TIVE DATA TO DEBUNK MYTHS ABOUT CHILDREN RAISED IN SAME-SEX-PARENTED FAMILIES

An issue that has received significant public and academic attention is how children raised in same-sex-parented families fare relative to other children. While a majority of scholars maintain that same-sex couples provide children with nurturing home environments where they can flourish (see e.g., Biblarz and Stacey 2010; Rosenfeld 2010), a loud minority has questioned same-sex parents' ability to raise well-adjusted children (see e.g., Marks 2012; Regnerus 2012; Allen 2015). More than 80 empirical studies across the social sciences have attempted to examine this issue (Center for the Study of Inequality 2017). Most of these studies were based on analyses of social surveys and concluded that there is no fundamental disadvantage associated

with being raised by same-sex parents (Center for the Study of Inequality 2017). The relevance of these findings is not merely academic, as demonstrated by their use in legislative processes. For example, this evidence was sought by the United States Supreme Court to challenge federal definitions that limited marriage to different-sex couples in *Obergefell vs Hodges 2015* (Manning, Fettro and Lamidi 2014; Gates 2015).

However, multiple voices have questioned the robustness of the evidence base on methodological grounds, claiming that existing studies rely on substandard data (Marks 2012; Regnerus, 2012; Allen 2015). Critics have argued that such data are generally not representative of the general population (non-probability sampling is the norm), encompass very few children in same-sex-parented families (samples of ∼30-60 children are common), rely on parent-reported assessments of how well their children are developing (which may introduce bias), and do not track children since birth. The latter is important, as many children enter same-sex couples after the dissolution of a different-sex-parented family. As such, statistical adjustments are required to ensure that any negative outcomes on children stemming from family dissolution are not 'blamed' on their subsequent families (Rosenfeld 2010).

This was the state of the literature when my co-author and I found a way to overcome these shortcomings by deploying administrative data from the Netherlands (Kabátek and Perales 2020). The Netherlands is internationally distinctive for its commitment to making administrative data accessible to social researchers through secure protocols and a strict vetting system. Through these means, we were able to access and analyze linked data from multiple registers capturing rich information on children and their families over the 2006-2018 period. This included information on family type (same-sex-parented or different-sex-parented),

demographic characteristics and socio-economic status, and children's educational outcomes and lifetime family transitions.

These data allowed us to resolve the methodological issues plaguing previous studies: they captured the whole population of children in the Netherlands (∼1.5 million children) including more than 3,000 children in same-sex-parented families, they incorporated objective and verifiable information on children's developmental outcomes in the domain of education (e.g., official records of school progress and test scores), and they tracked children's circumstances since birth. Further, the fact that the data originated in the Netherlands—a particularly progressive country concerning public opinion and legislation on same-sex relations—minimized the risk of confounding due to stigma and discrimination.

Using these unique and novel—yet readily available—dataset, we were able to provide unprecedentedly robust answers to the long-standing question of how children in same-sex-parented families fare. Our findings directly contradicted deficit models of same-sex-parenting: all else being equal, children raised by same-sex couples outperformed children raised by different-sex couples on multiple indicators of academic performance—including standardized tests scores, high school graduation rates, and college enrolment rates (Kabátek and Perales 2020). Representative results are presented in Figure 2.

This case study clearly showcases how existing administrative datasets are sometimes better placed than traditional data sources—such as social surveys—to generate robust empirical evidence on timely social issues of legislative, public and scholarly relevance. It also demonstrates the value of investing in a strong national system that allows researchers to access linked administrative data.
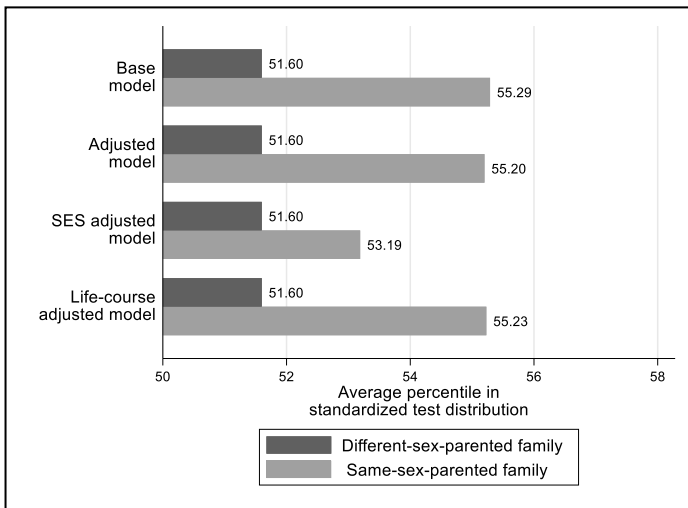
FIGURE 2: *Predicted standardized test scores of children raised in same-sex-parented and different-sex-parented families in the Netherlands*

Notes: Based on analyses in Kabátek & Perales (2020). The outcome variable is the overall score from a standardized test (*Cito toets*) taken by Dutch children at the end of primary school (Grade 8, ages 11-12). The test assesses children's capabilities in Dutch language, mathematics, and learning ability. The 'base model' is an unadjusted model. The 'adjusted model' controls for parental immigration, ethnic background, marital status and birth cohort, household residential characteristics, and child's sex and adoptee status. The 'socio-economic status (SES) adjusted model' adds controls for parental education, employment, disability, and income. The 'life-course adjusted model' adds controls for the number of residential moves since birth, the number of changes in family structure since birth, and the household structure at birth.

4.3 CASE STUDY 3: DEPLOYING AUSTRALIAN MATCHED EMPLOYER-EMPLOYEE DATA TO ASSESS THE EFFECTIVENESS OF LGBTIQ+ WORKPLACE INCLUSION POLICIES

Employers are progressively recognizing the importance of promoting workplace diversity and inclusion as a pathway to attracting and retaining talent, increasing productivity, and attaining a positive brand image (Badgett et al. 2013; Lloren and Parini 2017). Within the many faces of diversity, inclusion on the basis of sexual orientation remains comparatively under-researched (Ng and Rummens 2017). Theoretical perspectives grounded on the minority stress model point to complex relations between sexual-minority people,

co-workers, and employers. This is consistent with emerging empirical evidence indicating that non-heterosexual people face a range of unique workplace stressors, such as pressures to hide their identity at work, experiences of ostracism, jokes and innuendo targeting their sexuality, disproportionate rates of sexual harassment, and workplace policies that fail to incorporate their needs (Zurbrügg and Miner 2016; Lloren and Parini 2017; Ng and Rummens 2017).

In recent years, many employers in Australia and overseas have begun to invest significant resources towards the development of policies and practices aimed at redressing this situation. Popular initiatives include delivering diversity training to employees, promoting the formation and expansion of Ally networks, celebrating diversity in the workplace and local LGBTIQ+ role models, and identifying diversity champions amongst senior leadership (Lloren and Parini 2017; Pride in Diversity 2019). Despite these efforts, the degree to which inclusion policies achieve their intended aims of reducing workplace exclusion on the basis of sexual orientation and improving the work lives of sexual-minority employees remains an open question.

Unfortunately, Australia lacks an official, government-funded survey on workplace inclusion that could be deployed to empirically examine these issues. However, an Australian not-for-profit organization, ACON Health (ACON), has collected rich data in this space for over 10 years. ACON collects annual employer-level data on diversity practices to develop its Australian Workplace Equality Index (AWEI), a comprehensive and evidence-based measure gauging the extent to which employers' diversity practices align with international best-practice (Pride in Diversity 2019). To accompany the AWEI, ACON annually surveys employees from the participating organizations on their demographic characteristics, sexual orientation, gender identity, attitudes to workplace inclusion, and workplace experiences and behaviors (Pride in Diversity 2019). The AWEI Survey was issued for the first time in 2012 and has grown exponentially since then. In its 2020 iteration, it collected information from over 30,000 employees.

The original function of this data set was not for them to be analyzed by social researchers, but to support the implementation of the AWEI and the benchmarking of Australian workplaces in relation to their diversity practices. As such, they broadly fall within the realm of administrative data. Despite the critical importance of the AWEI data-collection exercises, ACON lacks the expertise to interrogate the data to their full potential. Within this context, a project co-lead between Dawn Hough (Director, Pride Inclusion Programs) and myself will establish a cross-sectoral collaboration to analyze ACON's data holdings using state-of-the-art quantitative methodologies, with the ultimate aim of improving our understanding of "what works" to enhance the inclusion of individuals from sexual minorities within Australian workplaces.

The projected program of research will involve linkages between the two types of data collected by ACON—the aggregate-level information on employer practices and the individual-level survey of employees' experiences—across multiple years. The employer-level data will offer detailed insights into the specific inclusion programs, policies and practices that each employer has put in place, and their quality (as evaluated by ACON). The individual-level data will bring in complementary insights into the degree to which sexual-minority individuals feel included, happy and productive in these workplaces. Combining these two sources of data into a multilevel, matched employer-employee dataset will enable us to explore the effects that various workplace inclusion policies have on individual employees. The logic of this approach is encapsulated in the schematic depicted in Figure 3. Empirical analyses of the resulting data will allow us to determine which policies work better across different settings, and to rank policies by their efficiency and ability to improve different outcomes for different types of workers. As a result, the project will yield concrete recommendations for employers and government alike.

This third case study showcases the value of administrative datasets held by non-government organizations, in this case a not-for-profit organization. As demonstrated here, these data can be creatively analyzed to generate unique insights that can fill significant gaps in scholarly and policy knowledge on sexual minorities. This case study also showcases the worth of cross-sectoral collaborations between academic and third-sector organizations as a means to improving equality of opportunity and outcomes by sexual orientation.
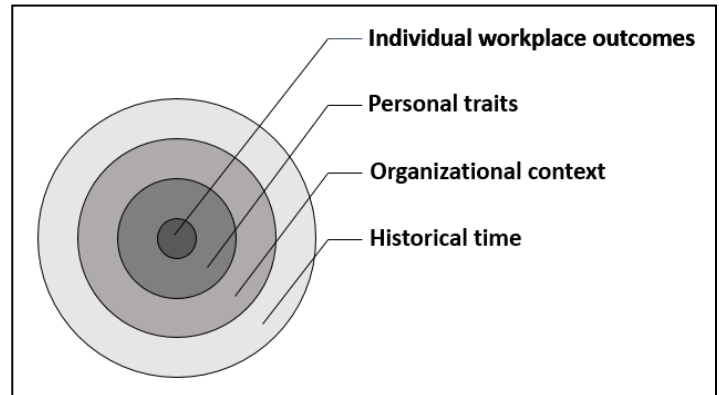
FIGURE 3: *A conceptual model of individuals' workplace outcomes*

5.   THE FUTURE OF ADMINISTRATIVE DATA TO EXAMINE INEQUALITY

This paper aimed at showcasing the benefits and opportunities associated with the use of administrative datasets to improve our understanding of processes with regard to equality of opportunity and outcomes. Collectively, the three case studies presented before exemplify how data sources that fall within the broad category of 'administrative data' can help social researchers generate new, policy-relevant knowledge on inequalities by sexual orientation. This knowledge is also important to contextualize public and legislative debates on timely social issues. Taken together, the case studies reflect several of the advantages of administrative datasets outlined in the earlier sections of the paper—such as their large sample sizes, which allow for detailed and robust analyses of small collectives, their provision of objective and novel measures, and their longitudinal and multilevel dimensions. Further, each of the case studies focused on a different life domain—health/wellbeing, family relations, and workplace inclusion—showcasing the versatility of these non-traditional data sources.

Despite the benefits of using administrative data for research aimed at understanding socio-economic inequalities, important obstacles also stand on the way. Here, I will outline three main challenges, acknowledging that other issues have also been discussed (see e.g., Australian Government Productivity Commission 2013; Connelly et al. 2016). First, administrative datasets do not always contain measures of key interest to inequality researchers. 'Blind spots' usually include variables capturing individuals' values, attitudes, aspirations and expectations, or subjective appraisals of their wellbeing (e.g., life satisfaction) and experiences (e.g., perceived discrimination). Under these circumstances, linkages of administrative and survey data constitute a valuable alternative to their separate use for research purposes. In the context of research on sexual orientation, traditional administrative data sources (e.g., Government records) usually fail to capture individuals' self-reported sexual identity, attractions or behavior, therefore restricting the analytic focus to individuals in officially recognized same-sex relations (Waite and Denier 2019). This also prevents a more nuanced separation of different sexual-minority groups (e.g., bisexual versus gay/lesbian people). Changes to the way in which government data are collected could help mitigate this limitation—for example, the inclusion of a sexual orientation question in the Census, as has been discussed in Australia (Parliament of Australia 2018).

Secondly, there are fundamental and legitimate concerns around individuals' right to privacy and the potential for harm (Harron et al. 2017). This includes individual harm through disclosure of sensitive information from their administrative records (Shepherd et al. 2019) and "networked harms" through the aggregate attribution of group-level characteristics to individuals (Boyd et al. 2014). Disclosure and privacy issues are particularly significant in research on sexual minorities, as sexuality is typically considered a highly personal realm and non-heterosexuality remains a stigmatized status. Administrative data use should be appropriately regulated to mitigate these risks. This requires a multi-pronged approach that incorporates: strong legislative frameworks, ethical scrutiny of projects using the data, de-identification strategies for administrative records,

processes for the screening and training of data users, secure channels for data analysis, and appropriate vetting of statistical outputs. The '5 safes' framework adopted by the ABS in recent years is an example of a successful approach (ABS 2017). Similar frameworks found in other countries—such as those implemented by the UK Office for National Statistics or Statistics Netherlands—have also proven to be effective in enabling the use of administrative data for research purposes while minimizing risk.

Third, it is also important to consider the 'social license to operate'. In most cases, it is not feasible to seek informed consent for administrative data use from all individuals for whom records are stored and shared. Because of this, the deployment of public collections of administrative data for research purposes should be built upon principles of public trust, and accompanied by efforts to ensure public understanding and awareness of how the data are being used and the benefits of doing so (see e.g., Sexton et al. 2017; Brown et al. 2019). Concerning sexual orientation, the public would benefit from accessible accounts of how studies utilizing administrative data have contributed to realizing equality of opportunity and outcome and lifting the life chances of sexual minority individuals. Once the social license is achieved, as I demonstrate in this article, there is substantial value in fully incorporating information on sexual orientation across standard administrative-data collections—including, but not limited to, the population Census.

Despite the need to navigate these significant challenges, it is safe to state that sociologists and social scientists working on issues of inequality have tremendous plans for administrative data. Based on the lessons learned from pioneering studies in this space, there are also great expectations about how the increasing availability and richness of these data will contribute to advancing the field (Connelly et al. 2016). Perhaps the most exciting prospect is that of steering towards solution-oriented social science research that can generate policy-relevant insights to redress socio-economic disparities between population groups (Australian Government Productivity Commission 2013; Card et al. 2010; Western 2019). The way I see it, administrative

data will play an important role in the future of inequality research.

6. REFERENCES

1. Allen, D. W. 2015. "More Heat than Light: A Critical Assessment of the Same-Sex Parenting Literature, 1995–2013." *Marriage and Family Review*, 51: 154-82.

2. Ambrey, C. L., Parsell, C., Spallek, M., and Robinson, R. N. 2019. "An Investigation into Repeat Requests for Charity: Evidence from the St Vincent de Paul Society Queensland, Australia." *Nonprofit and Voluntary Sector Quarterly*, 48 (1): 91-107.

3. Andersson, G., Noack, T., Seierstad, A., and Weedon-Fekjaer, H. 2006. "The Demographics of Same-Sex Marriages in Norway and Sweden." *Demography*, 43 (1): 79-98.

4. Australian Bureau of Statistics. 2017. "Managing the Risk of Disclosure: The Five Safes Framework (ABS Confidentiality Series, Aug 2017)." https://www.abs.gov.au/aus-stats/abs@.nsf/Latestprod-ucts/1160.0Main%20Features4Aug%202017. Accessed 12/08/20.

5. Australian Government Productivity Commission. 2013. "Using Administrative Data to Achieve Better Policy Outcomes". https://www.pc.gov.au/about/governance/an-nual-reports/2012-13/02-annual-report-2012-13-chapter1.pdf. Accessed 12/08/20.

6. Badgett, M. V., Choi, S. K., and Wilson, B. D. 2019. "LGBT Poverty in the United States: A Study of Differences between Sexual Orientation and Gender Identity Groups." UCLA, Williams Institute.

7. Badgett, M. V., Durso, L. E., Mallory, C., and Kastanis, A. 2013. "The Business Impact of LGBT-Supportive Workplace Policies." UCLA, Williams Institute.

8. Biblarz, T. J. and Stacey, J. 2010. "How Does the Gender of Parents Matter?" *Journal of Marriage and Family*, 72 (1): 3-22.

9. Boyd, D., Levy, K., and Marwick, A. 2014. "The Networked Nature of Algorithmic Discrimination." Data and Discrimination: Collected Essays. Open Technology Institute. https://www.danah.org/papers/2014/Data-Discrimination.pdf. Accessed 12/08/20.

10. Brown, A., Chouldechova, A., Putnam-Hornstein, E., Tobin, A., and Vaithianathan, R. 2019. "Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-Making in Child Welfare Services." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-12.

11. Bryman, A. 2016. *Social Research Methods.* Oxford, UK: Oxford University Press

12. Card, D., Chetty, R., Feldstein, M. S., and Saez, E. 2010. "Expanding Access to Administrative Data for Research in the United States." https://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf. Accessed 12/08/20.

13. Center for the Study of Inequality. 2017. "What Does the Scholarly Research Say about the Well-Being of Children with Gay or Lesbian Parents?" https://whatweknow.inequality.cornell.edu/top-ics/lgbt-equality/what-does-the-scholarly-re-search-say-about-the-wellbeing-of-children-with-gay-or-lesbian-parents/ Accessed 12/08/20.

14. Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." National Bureau of Economic Research, Working Paper 23618.

15. Connelly, R., Playford, C. J., Gayle, V., and Dibben, C. 2016. "The Role of Administrative Data in the Big Data Revolution in Social Science Research." *Social Science Research*, 59: 1-12.

16. DiPrete, T. A. and Eirich, G. M. 2006. "Cumulative Advantage as a Mechanism for Inequality: A Re

view of Theoretical and Empirical Developments." *Annual Review of Sociology*, 32: 271-97.

17. Drapeau, A., Boyer, R., and Diallo, F. B. 2011. "Discrepancies between Survey and Administrative Data on the Use of Mental Health Services in the General Population: Findings from a Study Conducted in Quebec." *BMC Public Health*, 11 (837): 1-10.

18. Drescher, J. 2015. "Out of DSM: Depathologizing Homosexuality." *Behavioral Sciences*, 5 (4): 565-75.

19. Einav, L. and Levin, J. 2014. "Economics in the Age of Big Data." *Science*, 346 (6210): 1-8.

20. Fraser, B., Pierse, N., Chisholm, E., and Cook, H. 2019. "LGBTIQ+ Homelessness: A Review of the Literature." *International Journal of Environmental Research and Public Health*, 16 (15): 2677.

21. Gates, G. J. 2015. "Marriage and Family: LGBT Individuals and Same-Sex Couples." *The Future of Children*, 25: 67-87.

22. Gravelle, T. B. and Carson, A. 2019. "Explaining the Australian Marriage Equality Vote: An Aggregate-Level Analysis." *Politics*, 39 (2): 186-201.

23. Harron, K., Dibben, C., Boyd, J., Hjern, A., Azimaee, M., Barreto, M. L., and Goldstein, H. 2017. "Challenges in Administrative Data Linkage for Research." *Big Data & Society*, 4 (2): 1-12.

24. Hatzenbuehler, M. L. 2009. "How Does Sexual Minority Stigma "Get Under the Skin"? A Psychological Mediation Framework." *Psychological Bulletin*, 135 (5): 707-30.

25. Institute of Medicine. 2011. "The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding." Washington, DC, USA: The National Academies Press.

26. Kabátek, J. and Perales, F. 2020. "Academic Achievement of Children in Same-Sex-Parented and Different-Sex-Parented Families: A Population-Level Analysis of Linked Administrative Data from the Netherlands." *Demography* 58 (2): 393-418.

27. Lloren, A. and Parini, L. 2017. "How LGBT-Supportive Workplace Policies Shape the Experience of Lesbian, Gay Men, and Bisexual Employees." *Sexuality Research & Social Policy*, 14 (3): 289-99.

28. Manning, W. D., Fettro, M. N., and Lamidi. E. 2014. "Child Well-Being in Same-Sex Parent Families: Review of Research Prepared for American Sociological Association Amicus Brief." *Population and Research Policy Review*, 33 (4): 485-502.

29. Marks, L. 2012. "Same-Sex Parenting and Children's Outcomes: A Closer Examination of the American Psychological Association's Brief on Lesbian and Gay Parenting." *Social Science Research*, 41: 735-51.

30. McCall, L. 2005. "The Complexity of Intersectionality." *Signs*, 30 (3): 1771-1800.

31. Mervis, J. 2017. "Data for All?" *Science,* 355 (6325): 573.

32. Meyer, I. H. 2003. "Prejudice, Social Stress, and Mental Health in Lesbian, Gay, and Bisexual Populations: Conceptual Issues and Research Evidence." *Psychological Bulletin*, 129 (5): 674-97.

33. Mullan, B. P. 2017. "The Sociology of Inequality and the Rise of Neo-Inequality." *Sociological Focus*, 50 (2): 105-124.

34. Ng, E. S. and Rumens, N. 2017. "Diversity and Inclusion for LGBT Workers: Current Issues and New Horizons for Research." *Canadian Journal of Administrative Sciences*, 34 (2): 109-20.

35. *Obergefell vs. Hodges*, 576 U.S. (2015) (no. 14-556), 2015 WL 2473451

36. Pachankis, J.E., Hatzenbuehler, M.L., Hickson, F., Weatherburn, P., Berg, R.C., Marcus, U., et al. 2015. "Hidden from Health: Structural Stigma, Sexual Orientation Concealment, and HIV across 38 Countries in the European MSM Internet Survey." *AIDS*, 29 (10): 1239-46.

37. Parliament of Australia. 2018. "The Question of Sexual Orientation... Something for Census 2021?" https://www.theguardian.com/australia-news/2021/aug/10/the-two-questions-likely-to-cause-the-most-controversy-on-australias-census-day Accessed 12/08/20.

38. Parsell, C., Petersen, M., and Culhane, D. 2017. "Cost Offsets of Supportive Housing: Evidence for

Social Work." *British Journal of Social Work*, 47 (5): 1534-53.

39. Perales, F. 2016. "The Costs of Being "Different": Sexual Identity and Subjective Wellbeing over the Life Course." *Social Indicators Research*, 127 (2): 827-49.

40. Perales, F. 2019. "The Health and Wellbeing of Australian Lesbian, Gay and Bisexual people: A Systematic Assessment Using a Longitudinal National Sample." *Australian and New Zealand Journal of Public Health*, 43 (3): 281-87.

41. Perales, F. and Campbell, A. 2018. "Who Supports Equal Rights for Same-Sex Couples? Evidence from Australia." *Family Matters*, 100: 28-41.

42. Perales, F. and Campbell, A. 2019. "Early Roots of Sexual-Orientation Health Disparities: Associations between Sexual Attraction, Health and Wellbeing in a National Sample of Australian Adolescents." *Journal of Epidemiology and Community Health*, 73 (10): 954-62.

43. Perales, F. and Todd, A. 2018. "Structural Stigma and the Health and Wellbeing of Australian LGB Populations: Exploiting Geographic Variation in the Results of the 2017 Same-Sex Marriage Plebiscite." *Social Science & Medicine*, 208 (July): 190-99.

44. Perales, F., Simpson Reeves, L., Plage, S. and Baxter, J. 2020. "The Family Lives of Australian Lesbian, Gay and Bisexual People: A Review of the Literature and a Research Agenda." *Sexuality Research & Social Policy*, 17 (1): 43-60.

45. Piketty, T. and Saez, E. 2014. "Inequality in the Long Run." *Science*, 344 (6186): 838-43.

46. Pride in Diversity. 2019. "Australian Workplace Equality Index 2019: The Definitive Benchmark on Australian LGBTIQ Workplace Inclusion." http://www.pid-awei.com.au/content/uploads/2015/08/AWEI Report 2019 Screen.pdf. Accessed 12/08/20.

47. Regnerus, M. 2012. "How Different are the Adult Children of Parents Who Have Same-Sex Relationships? Findings from the New Family Structures Study." *Social Science Research*, 41: 752-70.

48. Rosenfeld, M. 2010 "Nontraditional Families and Childhood Progress through School." *Demography*, 47 (3): 755-75.

49. Saxby, K., Sonja, C., and Petrie, D. 2020. "Structural Stigma and Sexual Orientation Disparities in Healthcare Use: Evidence from Australian Census-Linked-Administrative Data." *Social Science & Medicine*, 255 (June): 1-11.

50. Sexton, A., Shepherd, E., Duke-Williams, O., and Eveleigh, A. 2017. "A Balance of Trust in the Use of Government Administrative Data." *Archival Science*, 17 (4): 305-30.

51. Shepherd, E., Sexton, A., Duke-Williams, O., and Eveleigh, A. 2019. "Risk Identification and Management for the Research Use of Government Administrative Data." *Records Management Journal*, 30 (1): 101-23.

52. Short, M. E., Goetzel, R. Z., Pei, X., Tabrizi, M. J., Ozminkowski, R. J., Gibson, T. B., ... and Wilson, M. G. 2009. "How Accurate are Self-Reports? An Analysis of Self-Reported Healthcare Utilization and Absence when Compared to Administrative Data." *Journal of Occupational and Environmental Medicine*, 51 (7): 786-96.

53. Uhrig, N. 2015. "Sexual Orientation and Poverty in the UK: A Review and Top-Line Findings from the UK Household Longitudinal Study." *Journal of Research in Gender Studies*, 5 (1): 23–72.

54. Valfort, M. A. 2017. "LGBTI in OECD countries: A review." Paris: Organisation for Economic Cooperation and Development (OECD).

55. Waite, S. and Denier, N. 2019. "A Research Note on Canada's LGBT Data Landscape: Where We Are and What the Future Holds." *Canadian Review of Sociology*, 56 (1): 93-117.

56. Watts, D. J. 2017. "Should Social Science Be More Solution-Oriented?" *Nature Human Behaviour*, 1 (1): 1-5.

57. Western, M. 2019. "How to Increase the Relevance and Use of Social and Behavioral Science: Lessons for Policy-Makers, Researchers and Others." *Justice Evaluation Journal*, 2 (1): 18-34.

58. Wiik, K. A., Seierstad, A., and Noack, T. 2014. "Divorce in Norwegian Same-Sex Marriages and Registered Partnerships: The Role of Children." *Journal of Marriage and Family*, 76 (5): 919-29.

59. Wilson, T., Shalley, F. and Perales, F. 2020. "The Geography of Australia's Marriage Law Postal Survey Outcome." *Area*, 52 (10): 164-175.

60. Zurbrügg, L. and Miner, K. N. 2016. "Gender, Sexual Orientation, and Workplace Incivility: Who is Most Targeted and Who is Most Harmed?" *Frontiers in Psychology*, 7 (565): 1-12.