

## **Researching risk in the voluntary sector: The challenges and opportunities of regulatory data**

Diarmuid McDonnell and Alasdair C. Rutherford

### **Introduction**

The behaviour of voluntary organisations, and their willingness to be accountable, is a pressing policy issue around the world. In the UK for example, legitimacy and public trust are under threat due to a recent spate of high-profile voluntary sector crises and scandals, including concerns about large-scale and pervasive instances of financial mismanagement, intrusive and potentially harmful fundraising practices, and the abuse of vulnerable beneficiaries.

Concurrently, charity regulators are in a state of flux, dealing with declining or stagnating budgets and grappling with new strategic priorities to become data-led organisations.

Understanding the nature, extent and impact of risk is therefore of considerable importance for the field, sector, public, and policy practitioners.

This chapter reflects on the methodological implications and challenges associated with using regulatory data to study risk in the voluntary sector. In particular we describe collecting, operationalising and analysing the large-scale, often complex administrative data held by regulators that are necessary to study this topic. Drawing on numerous examples from a multi-year programme of research on the UK charity sector, we outline both the promise and the perils for researchers embarking on their own research.

### **The challenges and opportunities of regulatory data in measuring risk**

The nature of risk in the voluntary sector is broad and derived from the panoply of operational areas and decisions inherent in running organisations: ‘Financial, personnel, program and capital expenditure decisions all entail risk because they involve interactions with changing, complex, volatile or intrinsically stochastic economic, political and social environments.’ (Young, 2009: 33). In some areas our understanding of risk is good and backed by strong empirical evidence, in particular organisational failure (Helmig et al., 2014) and financial vulnerability (see Dayson, 2013). There are some where there is a burgeoning set of empirical work being conducted, like organisational fraud (Archambeault et al., 2015) and non-financial disclosures. But there remain topics that are under-theorised or lacking in robust or plentiful empirical examination, such as mission completion (Helmig et al., 2014), ethical fundraising practice, and safeguarding of vulnerable beneficiaries. What unifies many attempts at examining risk is the use of data held by bodies with responsibility for overseeing voluntary organisations, in particular the various charity regulators.

Data collected, processed and shared by charity regulators are classified as *administrative data*. Administrative data are often referred to as by-product data and are collected routinely by an organisation through discharging its functions: from a research perspective the data are ‘found’ rather than ‘made’ (Connolly et al., 2016). This type of data is collected by government departments and other organisations for the purposes of registration, transaction and record keeping, usually during the delivery of a service (Hand, 2018). The repurposing of administrative data for social science research is a long-standing challenge, with numerous large-scale initiatives seeking to support greater use of these data by the research community (e.g., Administrative Data Research Network in the UK). From an analytical perspective, administrative data can be used ‘to shed light on what has happened, to help to predict what

might happen in the future, and to evaluate systems and their performance, i.e. the data can later be subjected to statistical analysis.’ (Hand, 2018: 556)

It is our strong belief that grasping the ‘biography’ of administrative data is essential for producing analyses that minimise misunderstandings and mistakes (Foster et al., 2017; Hand, 2018). This applies whether you are a quantitative researcher building predictive models of a rare outcome in the voluntary sector (McDonnell & Rutherford, 2019), or a qualitative researcher using charity regulatory data as a sampling frame (Morgan & Fletcher, 2013). Thus, the focus of the rest of this chapter is on recognising and addressing some of the common issues that arise in the research use of administrative data. We do so by drawing on examples from our own programme of research on the voluntary sector in the UK and other English-speaking nations in the Global North.

Table 1 summarises some the main beneficial features of using administrative data for research purposes, as well as offering rejoinders to some of these claims (Hand, 2018; Connolly et al., 2016; Smith et al., 2004; Wallgren & Wallgren, 2007). The rest of this section delves into each of these features and relates them to study of risk in the voluntary sector.

**Table 1. Key features of administrative regulatory data**

Feature	Promise	Peril
Access	Lower cost to obtain compared to other forms of social data collection methods. Unobtrusive to units of analysis i.e., data are already collected.	Lots of effort to attain, clean and analyse; may have to pay for access. Serious concerns about informed consent i.e., data collected for some other non-research purpose.
Coverage	High/complete coverage of units of analysis, time periods, and/or variables.	Missing observations for units of analysis are common, as are missing variables.
Quality	Data are often high quality due to its role in functioning of organisation.	Organisational pressures / lack of expertise can produce low-quality data resources. Data collection process can change over time, overwriting or invalidating previous records or measures.
Capturing Reality	Closer to social reality than other data resources / collection methods; record <i>actual</i> events, behaviours and characteristics, not those that are <i>stated</i> or <i>claimed</i> .	Data are often still dependent on self-reporting e.g., self-assessment income tax returns.
Timely	Contemporaneous, continuous updating of fields.	There can be considerable lags between when data are generated, collected, processed and shared. Continuous collection can overwrite previous fields, leading to lost information.
Measurement	Tighter definitions for some measures e.g., defined in law.	Dependent on what is measured by data collector; data are not collected with important social concepts and measures in mind.
Granularity	Data often record information in considerable detail. Capture hard-to-reach samples / rare events / older time periods.	Spurious precision or unnecessary detail may hinder analysis of broader social concepts. Confidentiality concerns can arise with granular records or information.

---

Linkage	Data often contain unique ids or sufficient information for each unit of analysis to permit linkage to other sources of data.	Unique ids not always recorded diligently. Linking multiple, low-quality datasets compounds rather than alleviates issues.
---------	---	---

---

## *Access*

Charity regulators play an important role in public accountability and transparency for the voluntary sector. As a result, data on charities is made publicly available in many jurisdictions. This is a real benefit to researchers, as data can be collected without placing additional burden on participants. Of course, it comes with the constraints of working with any secondary data, that the researcher does not have control over the content and methods used to collect the data. But there is great potential for large amounts of detailed data to be freely accessible.

Case Study 1 – Charity data from the Charity Commission for England & Wales. The data collected on charities by the Charity Commission for England & Wales is freely available in an online searchable database (<https://register-of-charities.charitycommission.gov.uk/>). Details include contact details, finances, activities, trustees, governance policies and more. Many data items are available both contemporaneously and historically. Charities can be found via either their charity registration number or a keyword search. There is also a bulk data download of the current register available, including a selection of the available fields.

However, in many jurisdictions the rationale for making data available is to provide information about individual charities on the register to interested parties, rather than to make data available for research. This means that it can vary widely how straightforward it is to access and download register data at scale. Data may be available in a downloadable file, through an Application Programming Interface (API), or simply on a webpage. This can create significant barriers in either the technical knowledge or researcher time required to get access to even publicly available data. Some data may be available on request from the regulator, but even then the researcher may not be able to control the format in which the data is supplied.

## Case Study 2 – Accessing data on charity dissolution across international jurisdictions

(McDonnell, Rutherford & Cordery, forthcoming). Comparative work requires accessing and analysing data from across charity jurisdictions. While manual approaches suffice for some jurisdictions, it is necessary or more robust to use a suite of different computational methods of access: this is due to the use of differing data formats and web platforms in particular. In this project we created programming scripts to access and prepare data from eight jurisdictions, and made it available to researchers online: <https://github.com/DiarmuidM/charity-dissolution>.

Methods of data access are not necessarily stable. Data owners can and do change the mode of access, the layout of information, and the content shared – see for example the 2020 changes to the England and Wales Register of Charities. This makes it difficult to establish reliable methods of data access, and can jeopardise the building of consistent time series. Documenting and archiving data that is used for analysis is important to support the reproducibility of research.

### ***Coverage***

Administrative data is likely to feature good coverage of the population that it encompasses. For example, in Scotland an organisation is by definition a charity only if it appears on the Scottish Charity Register. Working with the data therefore means working with the population of charities, as defined. There are limits to this coverage however. A register may exclude organisations that meet the regulator’s criteria for inclusion (e.g. unregistered nonprofit organisations in Canada, or very small charities in England and Wales). Or it may include organisations that do not meet the researcher’s criteria (e.g. public universities, or private schools). So while coverage of register data is likely to be good, it is not within the researcher’s direct control, and one must understand the criteria for an organisation being part of a regulator’s register data.

The content of the data collected is driven by the legal context in which the regulator operates. For example, in Scotland the regulator has a focus on public trust in the charitable sector, and detailed data is collected on the governance of charities. In Canada, responsibility for charities sits within the tax authority, and extremely detailed and high-quality financial data is collected.

Missing data is a perennial challenge for researchers, and administrative data is no exception to this. Organisations are likely to have some sort of legal duty to report to a regulator, but the powers to enforce this and penalties for non-compliance vary widely across jurisdictions. There may also not be an obligation for organisations to complete all the fields in a form. There can be significant delays in data being reported and made available, as there is usually an allowance of several months after a reporting period ends before an organisation must complete a return. Care must be taken when using a time period where not all returns have been submitted as there could be a systematic pattern in the organisations that report early and those that report late. Regulators vary in their capacity to check information, and so individual data items can be missing, or ‘total’ fields may not equal the sum of their constituent parts. Researchers need to engage in significant data cleaning and quality-checking before conducting analysis on register data.

Case Study 3 – Examining charity complaints and regulatory investigations (McDonnell & Rutherford, 2018) The Scottish Charity Regulator provided a dataset of complaints about charities and the ensuing regulatory investigations or actions. This dataset has by definition complete coverage: organisations must be on the register to be considered charities, and it includes all complaints and investigations made by the regulator between 2006 and 2014.

However, there are some limitations in the information included in these records: researchers are reliant on the categorisation of complaints and investigations made by the regulator, rather than the full circumstances of the complaint. The nature of the data, and the level of detail, changes



over time. As a time series it is vulnerable to changes in regulatory practice that might change the meaning of particular terms, or lead to particular data items no longer being recorded.

### *Data Quality*

Administrative data are often vital to the effective functioning of an organisation. This is particularly true of charity regulators, who in many cases have a legal duty to maintain an accurate, public register of charities in their jurisdiction. Thus, it seems a reasonable assumption that data collected and shared by charity regulators will be of sufficient quality to permit robust analyses of the voluntary sector. One such example would be the financial information submitted by these organisations to a charity regulator. The legal status of these returns may lead us to expect that the quality of the data is high. But the data is collected for reporting and regulatory purposes, and not primarily for research. As such, care must be taken to understand how questions have been asked each year, and to account for changes in the form of data collection, the definitions of key terms, or the intended purpose of the data collection that might influence how and what is reported by organisations to the regulator.

We might expect that financial data is of a high quality, and for larger organisations this is likely true, particularly where accounts have had to undergo an external audit or inspection process.

But other data items, such as staff or volunteer numbers, can be of more variable quality particularly where there are not tight definitions of how to count or record this information.

Unlike with primary longitudinal survey data there is rarely an obligation to maintain consistent question wording or collection methods over time, with returns instead designed to collect the information required to satisfy legal and regulatory requirements at a given time point. So quality of a data item may vary over time; the definition of a key term may change; or a series of questions may be discontinued or reworded.

Case Study 4 – Changing questions, moving measures (Pennerstorfer & Rutherford, 2019) The Scottish Charity Regulator collects data on ‘Net Current Assets’ and ‘Total Funds’ to capture the assets held by Scottish charities. In 2011 a small change to question wording in the assets section of the annual return form led a small number of charitable trusts with significant assets to reclassify the reporting of their assets out of the ‘Net Current Assets’ measure while still including them in the ‘Total Funds’ measure. The sums of money involved were large enough to move the sector wide aggregate total of Net Current Assets down without any actual change in the underlying sector’s financial position. This highlights the importance of closely checking the annual return forms and being aware of even seemingly minor changes to wording or structure.

### *Measurement*

The use of administrative data provides a degree of measurement validity and accuracy, due to the tight, legal definitions of certain organisational events, statuses and characteristics. For example, researchers interested in examining charity dissolution in England and Wales using regulatory data can do so confident in the knowledge that the relevant categories in the data – ‘ceased to exist’ and ‘does not operate’ – are legally defined under section 34 of the Charities Act 2011. Similarly, the preparation of annual accounts and reports must or should follow a set of standards and rules (known as Statement of Recommended Practice [SORP] in the UK), resulting in a higher degree of reliability and validity across a range of financial metrics (e.g., the level of unrestricted reserves held by a charity).

However, even if the data can be considered as high quality (e.g., complete coverage of population, well-structured, granular, or accurate), measurement issues may still bedevil or preclude robust analysis of the data. This is due to a fundamental but understandable reality: administrative data are not collected with important social concepts and measures in mind.

Therefore, there can be little overlap between what is measured in the data and the information you need to examine the social concept or phenomenon of interest.

The measurement of risk is particularly dependent on high quality measurement. Despite the reams of information charity regulators collect on the voluntary organisations they oversee, identifying predictive indicators of important risks remains challenging. This can be seen in the reasonably frequent changes that are made to the variety of mandatory and voluntary reporting mechanisms operated by charity regulators. For example, in 2008 the Internal Revenue Service (IRS) radically redesigned the 990 form to capture information relating to governance and other non-financial domains of US nonprofits (Brody, 2012a). This was an attempt to uncover undesirable events and behaviours that previously lay hidden, such as financial relationships or familial relationships between trustees. However, to date there is no strong evidence that capturing information on these domains has led to improved governance in the US nonprofit sector (Brody, 2012b).

Case Study 5 – Improving charity accountability (McDonnell, 2017) The Scottish Charity Regulator previously collected data on a number of ‘accountability issues’ that the regulator believed were indicative of improper or concerning behaviour by charities. Many of these indicators focused on financial concerns, such as the possible misuse of charitable funds, or the use of unauthorised fundraising methods or third parties. There were 33 of these indicators, which were captured on an annual basis by the regulator. Analysis of these indicators revealed that a majority of Scottish charities included in the study trigger at least one of these accountability concerns and a minority do so persistently; however, no link was found between the indicators and ‘more objective’ negative organizational outcomes such as being reported to the regulator for alleged poor behaviour, regulatory intervention or charity dissolution.

## *Capturing Reality*

Administrative data is often considered to more closely capture aspects of social reality than other data collection methods like social surveys or participant observation (Hand, 2018). Such a claim is founded on the ‘objective’ process by which the information is generated or captured: that is, produced during the operation of some administrative function rather than a self-reporting mechanism. Consider measuring the risk of fraud in the voluntary sector: it seems self-evident that complaints and casework data from charity regulators would better capture the nature and prevalence of this phenomenon than a social survey reliant on sufficient participation and honesty from voluntary organisations (see case study 3).

However this paints a misleading picture, as many fields in administrative datasets rely on self-reported information. Regulatory data on voluntary organisations suffer from the same issue, where even key financial information is produced and reported in an inconsistent manner. For example, understanding the level of reserves a charity possesses - and thus gaining a sense of its resilience in the face of financial shocks - can only be reliably measured for a small subgroup of organisations in the UK (those generating over £500,000 in annual gross income), and even then consistency is not guaranteed due to differences in the interpretation of financial standards and guidance (Morgan, 2011).

## *Timely*

Charity regulation is a continuous process, and a significant advantage of using administrative data is the potential for the data to be ‘real time’ rather than requiring lengthy fieldwork.

Regulators usually collect annual returns from charities, which often includes their annual financial reporting. However, there are two main threats to having timely access to

administrative data. Firstly, there is usually a lag in organisations having to report to the regulator. This could be as much as a year beyond the data of the relevant financial period. Secondly, there can be a delay in the regulator's release of updated data, typically anything between 'real time' and quarterly.

The lag in charity reporting places a constraint on how recent the information available can be. If an organisation is allowed six months following their reporting period to submit their return, then the data supplied will concern a 12 month period that runs between 18 and 6 months previously. But organisations may also submit with different delays; some submitting early while others use the whole period, or submit late. This means that care must be taken in analysing data received within this 'submission window' as it will show a partial, and perhaps biased, picture of the sector. Only once the submission deadline has passed can the data be considered complete.

The delay in a regulator releasing information reflects their processes for administering the returns data, and their methods of data sharing. It is important to understand this schedule, and the implications it has for the data collected. A regular schedule for a defined period (e.g. quarterly) provides clarity for data availability, albeit with some delay. 'Real time' updating reduces the delay, but care must be taken before concluding that a given period's data collection is 'complete'. It is also important to be aware of any of seasonal variations, such as spikes in reporting around financial/tax year end; or annual 'bulk updates' carried out by the regulator such as removing dormant organisations.

Finally, 'timely' also reflects the fact that register data is usually a 'live' database. That means that new data may overwrite old data. For example, if a charity changes its areas of activity, these will be updated in the register. The old activities will no longer be listed, and there may not even be a public record that these have been changed. This means that contemporaneous data

may differ from an earlier extract, which can create some challenges for replications unless historic data is stored and well-documented.

Case Study 6 – Capturing the impact of Covid-19 on the voluntary sector (McDonnell & Rutherford, 2020b). Covid-19 represents an existential threat to many voluntary organisations, while simultaneously spurring new, large-scale forms of voluntary activity. Using comprehensive publicly available data from seven jurisdictions, we examined the impact of Covid-19 on the foundation and dissolution of charitable organisations. We employed an ‘excess events’ analytical approach, comparing the numbers of foundations and dissolutions in 2020 to what we would expect based on the trends from previous years. What is obvious from collecting and analysing the data is that the ‘live’ public registers are actually a work-in-progress: the work of regulators was disrupted by the pandemic (particularly during the first half of 2020), leading to delays in the registration and removal of organisations from the registers. This considerably hampers efforts to track the ‘real time’ impact of the public health emergency on the voluntary sector.

### ***Granularity***

The large sample sizes, good coverage and availability of historical data together mean that regulatory data can provide good granularity, and opportunities to study tightly defined sub-groups within the voluntary sector.

Researchers are dependent on the level of granularity available to them in the data supplied by a regulator, whether it is made available publicly or confidentially for research purposes. Often details will be categorised for recording, and cases may also be aggregated or anonymised in a way that makes it difficult to explore patterns at the level of the individual organisation.

Case Study 7 – Aggregate versus individual-level data: studying notifiable events in UK charities (McDonnell & Rutherford, 2019; McDonnell & Rutherford, 2020a). Charities in the UK are expected to notify their respective regulators of serious incidents that materially affect the charity. While the details of these schemes vary across jurisdictions, they collect similar information about risk voluntarily reported to regulators. In Scotland, we accessed charity-level information about the reporting of notifiable events that allowed us to explore which sorts of charities were most likely to report, and the sorts of concerns that they highlighted. However, for England & Wales researchers are only able to access aggregate monthly statistics for notifiable events. This allowed us to study patterns across the sector in the level of reporting in response to charity scandals, but not to examine these patterns at the level of the charity itself.

Events that are rare, and so difficult to capture in a random sample, can be filtered on to create a sub-sample for analysis with only the organisations of interest. Where historical data is available, these criteria can be applied over long periods in order to increase the size of the relevant group. A threat to this granularity is the potential for rare events to be sensitive, and to risk publicly identifying the organisation involved. This would be a concern when working with data on complaints, investigations or serious incidents, and particularly when using regulatory data that is not normally shared publicly.

Case Study 8 – Working with data on rare events (McDonnell & Rutherford, 2019). The numbers of charities voluntarily reporting serious incidents to the regulator is low in Scotland. Working with this data allows us to explore what sorts of organisations are proactively identifying risks and reporting them to the regulator. We can explore in detail the characteristics of this small group, and the regulator's ensuing actions. However this data is not publicly available, and details of some of the incidents are sensitive. Care had to be taken in analysing

and reporting on this data not to disclose the identity of individual organisations, particularly where they had recognisable characteristics.

## **Linkage**

Administrative data on organisations can provide opportunities to link data to other sources. Data on organisations may be held by other regulators, membership or infrastructure bodies. For example, in the UK many charities are also registered companies and must file returns and accounts with Companies House. Linking these datasets provides a richer picture of the operation of organisations. But a barrier to this linkage is the use of consistent unique organisational identifiers. While regulators often use their own unique identifiers to identify organisations, they do not necessarily hold the corresponding identifier for another regulator on file. In England and Wales, the Charity Commission does also record company numbers, allowing charities to be linked to their corresponding company record. However in Scotland the regulator does not record these identifiers, meaning that linkage must be done more crudely based on matches in name and address.

Case Study 9 – Linking Organisational Data to Charity Register Data (Brook & Rutherford, 2017) Data on third sector organisations who have interacted with their local Third Sector Interface (TSI) in Scotland were gathered to explore the distribution of organisations across the country. For some organisations a charity number had been recorded. This permits an unambiguous one-to-one linkage between the database record and a Scottish charity. For others, no identifier was collected, and organisation records could be probabilistically matched between the database and the charity register on organisation name and address. For many, however, they were voluntary organisations that are not registered charities, and so no record of them is kept by the charity regulator.



## **Reflections on using regulatory data to measure risk**

We argue that there is great potential to use regulatory data to better understand risk in the voluntary sector. The intersection of several distinct developments - increased data availability, technology to gather and analyse this data, and interest in comparative research - have created real opportunities to address a range of important research questions at scale. A desire from many regulators to increase transparency, and a focus on public trust in the voluntary sector, has spurred developments to share more detailed data online. The keen interest of many charity and nonprofit regulators in a risk-based approach mean that routine data is available to consider risk beyond the financial domain, and across many dimensions.

But we also caution that easy access to data is not a panacea for the analysis of risk. Even when data is 'open' and 'publicly available' there can still be significant technical challenges in accessing and working with it, both through the methods that must be used (e.g., programming scripts) and the file formats or data structures in which information is shared (e.g., JSON).

While it is easy to focus on the technical challenge of data access, it is just as important that regulatory data be viewed through the lens of critical social science. There is a high bar in the institutional, cultural and contextual knowledge required to make sense of the data that is available, and significant diversity across jurisdictions in the meaning and form of similar phenomena. As social scientists, we still need to consider issues in the measurement, operationalisation and power dynamics of working with regulatory data. Some information is provided by charities as a legal duty, while other items are voluntary reports. Data collected with the intention of evaluating or 'policing' organisations creates a set of incentives for both the regulator and charities in how organisations and their activities are presented.

Our discussion here has primarily considered charity data held in English-speaking ‘Global North’ countries. Many of the challenges we discuss will also apply to regulatory data in other countries. In many jurisdictions, however, this sort of regulatory data is not collected (e.g., Ghana); is collected but not shared (e.g., China, Japan); or is publicly available but not in digital machine-readable formats (e.g., Singapore, India). This creates additional barriers to accessing data in the first instance, as well as ongoing cultural, linguistic and substantive difficulties in parsing and analyzing the data.

But despite these challenges, regulatory data does provide the opportunity to gain real understanding of risks within the voluntary sector at a scale that would be infeasible with primary data collection. Insights from this data have the potential to support the evaluation of public policies, and help to build public trust in the voluntary sector. Regulators have the opportunity to better target their data collection, and increase the ease of access. Voluntary organisations can benefit from understanding the determinants and predictors of risk across the sector. So finally, we would like to call for greater collaboration between researchers, regulators, and voluntary organisations to collect and share high quality data that meets all these needs. With suitable consideration of the perils, there is significant promise in the analysis of regulatory data on voluntary organisations globally.

### **Further readings**

Elizabeth Bloodgood and colleagues’ (2014) work on national variation in nonprofit regulation provides an excellent account of the political and legal context these organisations operate in globally.

To understand recent shifts in the resourcing and focus of charity regulators, we highly recommend Oonagh Breen's 2018 overview.

Brooker's (2020) guide to computational approaches for collecting and processing social data is a useful starting point for researchers interested in developing their programming knowledge and skills.

*Society Organisations in Scotland*. Working Paper. **Bibliography** Archambeault, D. S., Webber, S., & Greenlee, J. (2015) 'Fraud and Corruption in U.S. Nonprofit Entities: A Summary of Press Reports 2008-2011', *Nonprofit and Voluntary Sector Quarterly*, 44(6): 1194-1224.

Bloodgood, E. A., Tremblay-Boire, J., & Prakash, A. (2014) 'National styles of NGO regulation', *Nonprofit and Voluntary Sector Quarterly*, 43(4): 716-736.

Breen, O. (2018) 'Redefining the measure of success: a historical and comparative look at charity regulation', in M. Harding (ed.), *Research Handbook on Not-For-Profit Law*, Cheltenham: Edward Elgar Publishing Limited. pp. 549–569.

Brody, E. (2012a) 'U.S. Nonprofit Law Reform: The Role of Private Organizations', *Nonprofit and Voluntary Sector Quarterly*, 41(4): 535-559.

Brody, E. (2012b) 'Sunshine and Shadows on Charity Governance: Public Disclosure as a Regulatory Tool', *Florida Tax Review*, 12(4): 183-234.

Brook, O., & Rutherford, A. C. (2017) *Using Administrative Data to Understand Civil*

Brooker, P. (2020) *Programming with Python for Social Scientists*. London: SAGE Publications Ltd.

Connolly, 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.

Dayson, C. (2013) 'Understanding financial vulnerability in UK third sector organisations: methodological considerations and applications for policy, practice and research', *Voluntary Sector Review*, 4(1): 19-38.

Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., & Lane, J. (2017) 'Introduction', in I. Foster, R. Ghani, R. S. Jarmin, F. Kreuter, & J. Lane (eds.), *Big Data and Social Science: A Practical Guide to Methods and Tools*, Boca Raton: CRC Press. pp. 1-22.

Hand, D. J. (2018) 'Statistical challenges of administrative and transaction data', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(3): 555-605.

Helmig, B., Ingerfurth, S., & Pinz, A. (2014) 'Success and Failure of Nonprofit Organizations: Theoretical Foundations, Empirical Evidence, and Future Research', *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 25(6): 1509-1538.

McDonnell, D. (2017) 'Improving charity accountability: Lessons from the Scottish experience', *Nonprofit and Voluntary Sector Quarterly*, 46(4): 725-746.

McDonnell, D., & Rutherford, A. C. (2018) 'The determinants of charity misconduct', *Nonprofit and Voluntary Sector Quarterly*, 47(1): 107-125.

McDonnell, D., & Rutherford, A. C. (2019) 'Promoting charity accountability: understanding disclosure of serious incidents', *Accounting Forum*, 43(1): 42-61.

McDonnell, D., & Rutherford, A. C. (2020a) *Incentivising charity accountability: Evidence from the Oxfam scandal*. Working Paper.

McDonnell, D., & Rutherford, A. C. (2020b) *The impact of Covid-19 on the foundation and dissolution of charitable organisations*, <https://diarmuidm.github.io/charity-covid19/index>.

McDonnell, D., & Rutherford, A. C., & Cordery, C. (forthcoming) 'Mission accomplished? A cross-national examination of charity dissolution', *Voluntas: International Journal of Voluntary and Nonprofit Organizations*.

Morgan, G. G. (2011) 'The use of UK charity accounts data for researching the performance of voluntary organisations', *Voluntary Sector Review*, 2(2): 213-230.

Morgan, G. G., & Fletcher, N. J. (2013) 'Mandatory Public Benefit Reporting as a Basis for Charity Accountability: Findings from England and Wales', *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 24(3): 805-830.

Pennerstorfer, A., & Rutherford, A. C. (2019) 'Measuring Growth of the Nonprofit Sector: The Choice of Indicator Matters', *Nonprofit and Voluntary Sector Quarterly*, 48(2): 440-456.

Smith, G., Noble, M., Anttilla, C., Gill, L., Zaidi, A., Wright, G., Dibben, C., & Barnes, H. (2004) *The Value of Linked Administrative Records for Longitudinal Analysis, Report to the ESRC National Longitudinal Strategy Committee*, Swindon: ESRC.

Wallgren, A., & Wallgren, B. (2007) *Register-based Statistics: Administrative Data for Statistical Purposes*, New York: John Wiley & Sons.

Young, D. R. (2009) 'How Nonprofit Organizations Manage Risk', in S. Destefanis & M. Musella (eds.), *Paid and Unpaid Labour in the Social Economy: An International Perspective*, (Heidelberg: Physica-Verlag. pp. 33-46.