Received: 11 January 2023 | Revised: 8 June 2023 | Accepted: 15 June 2023

DOI: 10.1111/puar.13693

RESEARCH ARTICLE

PAR PUBLIC ADMINISTRATION REVIEW ASPA

Fancy seeing you here…again: Uncovering individual-level panel data in repeated cross-sectional surveys

Benny Geys

Department of Economics, BI Norwegian Business School, Bergen, Norway

Correspondence
Benny Geys, Department of Economics, BI Norwegian Business School, Kong Christian Frederiks plass 5, 5006 Bergen, Norway.
Email: benny.geys@bi.no

Abstract

Many theories in Public Administration and Public Management explicitly relate to *changes over time* in the attitudes, values, perceptions, and/or motivations of public-sector employees. Examining such theories using (repeated) cross-sectional datasets may lead to biased inferences and an inability to expose credible causal relationships. As developing individual-level panel datasets is costly and time-consuming, this article presents a method to make better use of existing surveys fielded repeatedly among the same respondent pool without individual identifiers. Specifically, it sets out an approach to create a system of unique identifiers using information about respondents' background characteristics available within the original data. The result is a panel dataset that allows tracking (a subset of) individual respondents across time. The article discusses issues of feasibility, credibility as well as ethical considerations. The methodology has further practical value by highlighting data characteristics that can help minimize identifiability of respondents while creating public-release datasets.

Practitioner points

- Individual background characteristics offer possibilities to create individual-level panel data from repeated survey waves fielded among the same respondent pool.
- The resulting panel datasets allow improvements in the use and value of existing survey data for public sector management and practice.
- The presented methodology also offers a tool to help minimize identifiability of respondents when creating Public Release datasets.
- Ethical concerns regarding anonymity, identifiability, and purpose limitation can be handled with proper care and sensitivity by following well-designed procedures.

INTRODUCTION

Cross-sectional surveys involve the collection of information from a (possibly large) number of respondents at one single point in time. Examples of such data collection efforts in Public Administration scholarship include one-off surveys on job satisfaction or work–life balance among civil servants (Feeney & Stritch, 2019; Steijn & Van der Voet, 2019), on Organizational Citizenship Behavior in the public sector (de Geus et al., 2020; Vigoda-Gadot & Beeri, 2011), or on red tape perceptions among public employees (Hattke et al., 2018; Scott & Pandey, 2005). Cross-sectional analysis of such datasets has been critical to the development of many concepts and theories in Public Administration and Public Management (Stritch, 2017), and continues to be "a staple in [the] public administration methodological toolkit" (Pandey, 2017, p.135; see also Vigoda-Gadot & Vashdi, 2020). Stritch (2017), for instance, highlights that only about 10% of all 1041 studies published in top public administration journals between 2011 and 2015 were longitudinal in nature. Barely 10% of these longitudinal studies (i.e. 11 articles) had individuals as the unit of analysis. In similar vein, Ritz et al. (2016) illustrate that empirical studies on Public Service Motivation are mostly cross-sectional in nature (81.3%), while very few employ longitudinal data (7.4%).

This dearth of longitudinal public administration research *at the individual level* is at least in part due to data constraints. Developing panel datasets is costly and time-consuming. Furthermore, difficulties in repeatedly accessing large samples of active civil servants and restrictions imposed by data protection regulations add extra challenges. Data availability is, however, not the main bottleneck. In fact, many individual-level surveys among civil servants are fielded repeatedly in much the same format, often over long periods of time. The US Federal Employee Viewpoint Survey, for instance, is run on a (bi-)annual basis since 2002 (Fernandez et al., 2015), while the American State Administrators Project combines surveys targeting the universe of US state-level agency leaders every 4 or 6 years since 1964 (Yackee & Yackee, 2021). In similar vein, the Norwegian Administration Surveys are fielded every 10 years since 1976 among staff in Norwegian ministries and agencies (Christensen et al., 2018), the German Political-Administrative Elite survey every 4 years since 2005 among all senior civil servants in Germany (Ebinger et al., 2022) and Hussein Kassim and coauthors have run three large-scale surveys among European Commission staff in 2008, 2014, and 2018 (Kassim et al., 2013; Murdoch et al., 2021). Annual or multi-annual surveys among all civil service staff members have also been conducted in, for instance, Australia (Australian Public Service Employee Census), Canada (Canada Public Service Employee Survey), New Zealand (Te Taunaki Public Service Census), and the United Kingdom (UK Civil Service People Survey). Finally, many ministries, agencies, and other public sector organizations across a wide range of countries conduct regular internal surveys among their staff.

Unfortunately, many of these repeated cross-sectional surveys do not contain "a system of unique identifiers to track even a subsample of individual respondents over time" (Fernandez et al., 2015, p. 389; see also Bertelli et al., 2015; Stritch, 2017; Wynen et al., 2020). In other words, the information contained in many existing survey datasets is not immediately linkable at the individual level over time, which precludes the possibility to engage in longitudinal research *at the individual level*. This is particularly regrettable since longitudinal data offer far greater potential to derive credible causal inferences relative to cross-sectional data (King et al., 1994; Pandey, 2017; Stritch, 2017). They are also much better suited to gain insights into temporal dynamics in attitudes, values, perceptions, and/or motivations. Cross-sectional data are by their very nature static, and thus *cannot* reliably investigate the dynamics at the heart of many theories in Public Administration and Public Management, nor fully satisfy many public organizations' interest in "tracking progress over time" (paraphrased from the objectives of the 2022/23 Canadian Public Service Employee Survey).

This article proposes a simple approach to overcome this critical limitation of repeated cross-sectional surveys. Exploiting information about respondents' background characteristics available within the original datasets, it presents a method to create a system of unique identifiers for individual respondents in repeated cross-sectional surveys of the same target population. The result is a panel dataset at the individual level, which offers novel opportunities for the longitudinal analysis of a subset of respondents.[1] This not only provides great potential for moving forward our understanding of individual-level processes and temporal dynamics via the enhanced use of publicly available datasets. The same method can also be applied by practitioners on repeated cross-sectional surveys fielded internally within a country's civil service as well as individual ministries, agencies, or other public organizations. As such, it can improve the use and value of such datasets in informing public sector management and practice.

Creating unique identifiers for individual respondents based on detailed background characteristics naturally raises several pertinent ethical considerations with regard to respondent privacy, anonymity, consent, and confidentiality. Researchers implementing the proposed strategy should therefore carefully consider these aspects at every stage of the process, and the article discusses how these concerns can be addressed and handled with the utmost sensitivity. With such ethical considerations in mind, the methodology proposed also holds further practical value. That is, it draws attention to, and provides precise guidelines for, specific actions that can be undertaken—in terms of the re-coding and publication/withholding of specific background variables—to help minimize identifiability of respondents while creating Public Release datasets.

The next section sets out the proposed method, and discusses key issues of credibility, validity, feasibility, and applicability. Subsequently, we address potential ethical concerns and implications for respondents' informed consent, while the penultimate section briefly discusses a concrete example. The final section concludes with a discussion of strengths and limitations, while a practical implementation guide using *Stata* and *R* is included in Online Appendix A.

UNCOVERING INDIVIDUAL-LEVEL PANEL DATA FROM REPEATED CROSS-SECTIONS

When surveys are fielded as independent cross-sectional data collection efforts, they commonly do not include unique identifiers to link the data over time at the individual level (Fernandez et al., 2015; Stritch, 2017; Wynen et al., 2020). This precludes using the collected data for individual-level longitudinal analyses and the assessment of dynamic public administration theories. Yet, each wave of a given survey generally contains a set of socio-demographic characteristics, and thus collects—often

TABLE 1 Guidelines, rules of thumb and red flags.

	Rules of thumb and red flags
Discriminatory power	
Number of characteristics	A minimum of seven individual-level characteristics. Basic selection must include gender, age, organizational affiliation and time spent within the organization (or close equivalents). Operationalization of individual characteristics must involve at least five distinct values whenever possible and at least two characteristics must be measured using more than 20 distinct values.
Correlation between characteristics	Most individual-level characteristics must display pairwise correlation coefficients below 0.25 in absolute terms.
Stability of surveyed population	
Respondent selection strategy	Random and stratified random sampling should be avoided, while broad-based sampling strategies (e.g., complete census) are strongly preferred.
Staff turnover and data frequency	The longer the time period between survey waves, the lower the annual turnover of staff members should be.
Drop-out and repeat responders	
Response rate and sample size	Low response rates undermine the ability to observe respondents across survey waves. Response rates and required sample size are inversely related.
Representativeness and sensitivity	Researchers should *always* report any (im) balances in covariates at the initial point of measurement between respondents that can and cannot be tracked over time. Validation checks on the potential for false positives should be performed and reported (see Online Appendix B). False positives in excess of 5% of the final sample are a cause for concern.

quite detailed—information about the respondents. Many of these socio-demographic characteristics are (near-) fixed over time (such as gender, birth-year, birth-place, or nationality), develop at a fixed pace over time (such as age, or length of work experience), or remain part of an individual's history once they occur (such as specific career developments or obtaining an educational degree). These characteristics therefore represent something tangible about any given respondent, which we can expect to observe in (near-)identical fashion whenever this respondent participates in distinct waves of a survey. We maintain that a careful combination of such characteristics into detailed respondent profiles can be used to assess whether the same individuals re-appear in multiple waves of the same survey among the same target population.

Why can this work? As is well-known from probability theory, many combinations of background characteristics are extremely unlikely to appear more than once in any dataset whenever one considers a sufficiently large and detailed set of characteristics (Wendl, 2003). Observing this exact "profile" of characteristics therefore would pinpoint a specific respondent independent from any others. Moreover, detecting this profile across two consecutive survey waves among the same respondent population would make it very likely that we are dealing with the same respondent at both points in time (this argument is developed in more detail below). We maintain that this makes it possible to exploit detailed background information included in most cross-sectional surveys in order to extract a panel dataset containing (a subset of) respondents at several points in time. The remainder of this section discusses the credibility, validity, feasibility, applicability, and novelty of this approach, which culminates in a number of practical guidelines set out in Table 1. Further details on the derivation of these guidelines using simulations based on *both* real-world *and* computer-generated data are provided in Online Appendix B.

DISCRIMINATORY POWER

A first potential hurdle relates to a lack of discriminatory power in the available individual-level background characteristics. That is, many people are born in the same year (and thus have the same age), hold a degree in the same subject, grew up in the same municipality, or started working in their organization at the same point in time. Moreover, several individuals may well share a number of the same background characteristics, as there are many people born in, say, 1977 who hold a degree in economics and are male. This observation implies that certain combinations of background characteristics may not be able to isolate a unique respondent. Clearly, however, this holds only when multiple individuals perfectly share *all available background characteristics simultaneously*, which, in practice, becomes highly unlikely with a large enough number of sufficiently detailed background characteristics (these thresholds are further specified below based on various simulation exercises).

We can illustrate this point using a canonical example from probability theory (McKinney, 1966; Wendl, 2003). The so-called birthday paradox considers the probability that at least two people in a party of n randomly chosen individuals share the same birthday (referred to as a "collision"). This initially appears very improbable in small parties. Yet, this probability exceeds 50% already among a group of just 23 party-goers (McKinney, 1966; Wendl, 2003). The reason is that there are $22 \times 23 = 253$ pairs of birthdays to consider, which is well over half the number of days in a year. Crucially, however, the required number of party-goers increases rapidly when adding extra characteristics (e.g., gender or birthplace), and allowing for a non-uniform distribution of these characteristics (e.g., more men than women). Specifically, to find a man and a woman sharing the same birthday with more than 50% probability, there should be 32 party-goers if both groups are equally large (i.e. 16 men and 16 women), but 49 party-goers under a very unequal gender distribution (e.g., 43 men and six women). Adding more background characteristics or more detailed measurement of each characteristic (e.g., birthplace at the municipal rather than regional level) thus can quickly increase discriminatory power by making "collisions" extremely unlikely even in populations of the size relevant to public sector organizations (for a formal derivation, see Wendl, 2003). If a collision nonetheless occurs across two waves of the same survey, it is therefore exceedingly likely to represent a very specific respondent who reappears on both occasions.

An example based on the Norwegian Administration Surveys collected over the 1976–2016 period (Christensen et al., 2018) serves to illustrate this point further. Imagine the probability of observing a female, 28-year old respondent born in Oslo (the Norwegian capital), who studied law, did part of her degree abroad, speaks *Bokmål* (the most common language variant used in Norway), started working 6 years ago directly after finishing college, is member of a union and former member of a political party, never stood for election, and whose father only finished high-school and was a farmer. This is the level of detail available in the background characteristics included in the Norwegian Administration Surveys. The probability of observing such an (imaginary) respondent profile—given the actual distribution of these background characteristics in the dataset—is below 0.00000017 (or one in 5.9 million). With 3000 to 5000 individuals working in Norwegian ministries over the time period of these surveys, this makes it extremely unlikely to observe a respondent with the described profile once—let alone twice in two consecutive waves of the survey. When this nonetheless happens, it is most likely the same respondent.

Clearly, this simple calculation assumes that all background characteristics are independent of each other, which may not be realistic. Age and experience within an organization, for instance, are often strongly positively correlated (Pearson r equal to 0.61 and 0.36 in the Norwegian Administration Surveys and the American State Administrators Project datasets, respectively). Nonetheless, many other background characteristics remain at best weakly correlated in most real-world datasets. This holds, for instance, for age and gender (Pearson r equal to −0.09 and −0.04 in the Norwegian Administration Surveys and American State Administrators Project datasets, respectively). Moreover, even if some background characteristics often appear together, individuals sharing these characteristics usually continue to vary along other dimensions. In the Norwegian Administration Surveys, for example, respondents between 45 and 54 years ($N = 1753$) differ substantially in terms of their father's profession, years of experience in their ministry, and their municipality of birth (see top panel of Figure 1). The same is true for female respondents who hold a law degree and speak *Bokmål* ($N = 554$) (see bottom panel of Figure 1).

Simulations suggest that including a minimum of seven to nine individual-level characteristics offers high discriminatory power, particularly when at least some characteristics are measured using more than 20 categories (see Online Appendix Tables B.1 and B.2). Under such conditions, we can correctly uncover up to 100% of respondents appearing in multiple survey waves based on their specific background profile, while simultaneously limiting the number of "false positives" (i.e. respondents in two time periods incorrectly designated as the same individual). Using fewer than seven background characteristics quickly decreases the number of correctly identified repeating respondents and increases the share of false positives, both of which undermine the validity of the resulting panel dataset (see Online Appendix Tables B.1 and B.2).[2] In public sector settings, gender, age, organizational affiliation and time spent within the organization (or close equivalents) arguably constitute "core" features of employees, and should be included in the basic selection of (seven or more) background variables. Ideally, age and time spent in the organization should thereby be measured in years rather than collected in bins (e.g., 5- or 10-year categories) to maximize the granularity and precision of the data (Table 1). Note also that the simulations in Online Appendix Table B.2 unsurprisingly highlight a rapidly decreasing discriminatory power when subsets of the variables included in the analysis are too strongly correlated (i.e. $r = .50$ or $r = .75$). In sharp contrast, the number of false positives and false negatives remains very low for weaker correlation levels (i.e. $r = .10$ and $r = .25$) as long as enough background characteristics are available (see above). Hence, in order to achieve sufficient validity and reliability of the resulting panel dataset, one can set a rule of thumb that most of the characteristics included in the analysis must display pairwise correlation coefficients below .25 in absolute terms (Table 1).[3]

These minimum requirements naturally affect the types of datasets where this approach can(not) be employed. The Federal Employee Viewpoint Survey, for instance, includes an insufficient number of background characteristics to allow for credible recovery of individual respondent profiles over time. In contrast, the publicly available 1964–2008 dataset of the American State Administrators Project (Yackee & Yackee, 2021) includes a sufficiently wide range of background characteristics to extract a viable panel dataset. In most cases, however, researchers will need to apply for restricted-use datasets.[4] For instance, the Public Release data files for the Australian Public Service Employee Census and the Norwegian Administration Surveys do not contain the necessary level of detail in terms of background

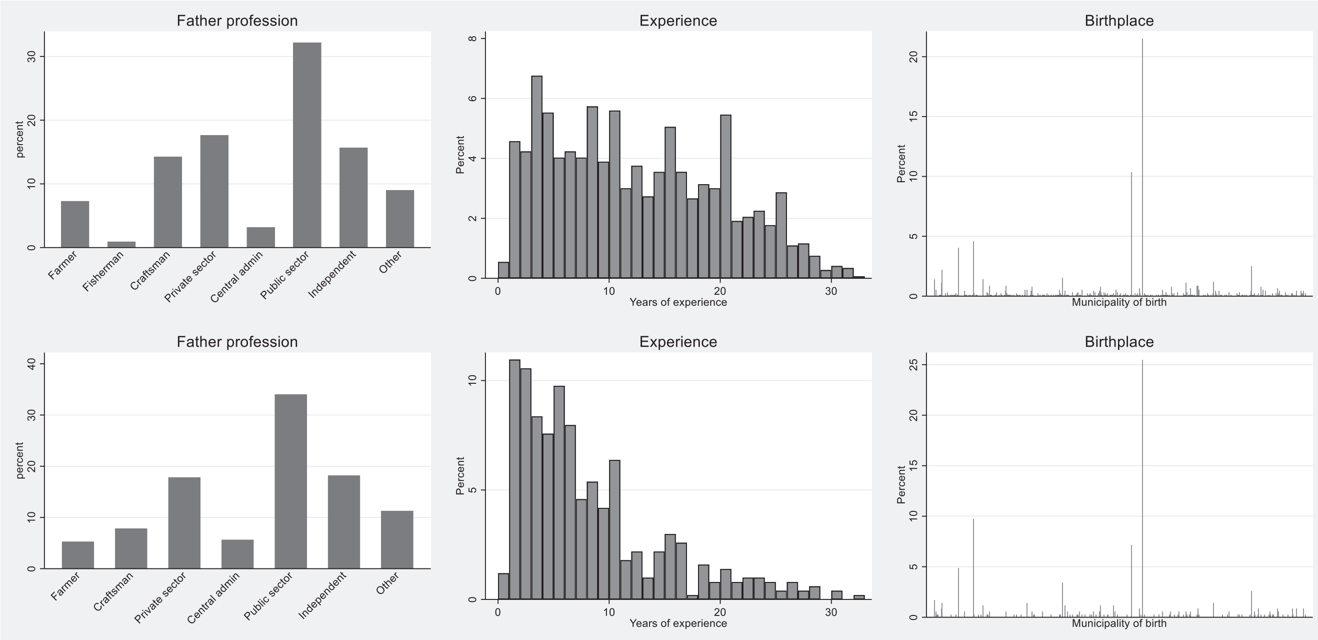

FIGURE 1 Distribution of background characteristic for respondents between 45 and 54 years (top row) or for women holding a law degree that speak Bokmål at home (bottom row). The figure includes only survey respondents between 45 and 54 years (top row), or women holding a law degree that speak Bokmål at home (bottom row). Each plot displays the distribution of three other background characteristics for this sample of respondents (i.e. father's profession, years of experience within their ministry, and municipality of birth).

characteristics, but the original datasets are much richer. The same also holds for the datasets from the German Political-Administrative Elite surveys (Ebinger et al., 2022), large-scale surveys among European Commission staff in 2008, 2014, and 2018 (Kassim et al., 2013; Murdoch et al., 2021), as well as numerous surveys fielded internally within ministries, agencies, or other public organizations. As such, the method proposed here has the potential for wide applicability (both by researchers and practitioners within public-sector organizations).

Stability of surveyed population

A second potential hurdle relates to a lack of stability in the surveyed population. If there is a low probability that the same person enters the respondent pool of the same survey on multiple occasions, there will be little overlap between the respondent samples across consecutive waves. Naturally, it then becomes extremely hard to observe responses for the same person more than once. This lack of stability in the surveyed population may arise for several reasons, some of which are under researchers' control while others are not.

A first reason is under the researcher's control as it relates to the sample selection strategy. Drawing a (stratified) random sample out of the targeted population is highly beneficial for many reasons, such as external validity and generalizability. However, it dramatically reduces the likelihood that the same person is drawn into the sample on multiple occasions—particularly when the targeted population is large and/or a small sample is surveyed—and a distinct group of people therefore enters consecutive data collection efforts. Reliance on (stratified) random sample selection strategies thus makes the method proposed in this article unfeasible. This excludes both population-wide surveys such as the World Values Survey or National Election Studies, and surveys among civil servants such as the US Federal Employee Viewpoint Survey.[5] Still, other surveys among public employees take a more general approach. The Norwegian Administration Surveys, for instance, target *all* ministerial staff with at least 1 year of service (Christensen et al., 2018), while the German Political-Administrative Elite surveys and American State Administrators Project surveys target *all* employees at the top hierarchical level(s) of the German national and US state-level bureaucracies (Ebinger et al., 2022; Yackee & Yackee, 2021). The Australian, Canadian, New Zealand, and UK public service employee censuses as well as the European Commission surveys mentioned above even address *all* staff members within these respective organizations. This broad-based approach strongly benefits the feasibility of extracting a panel dataset of individual respondents over multiple survey waves using individual-level background information.

A second reason for low stability in the respondent population lies beyond the control of researchers since it is related to staff turnover. High staff turnover limits the number of respondents in later waves of the survey who were employed in the same organization during a previous wave. Hence, uncovering the same respondents over multiple waves of the same survey requires a sufficiently stable workforce within the targeted organization(s). From this perspective, it is useful that many public sector organizations document low levels of employee turnover. Official European Commission statistics, for instance, show staff exits below 1% on an annual basis, as well as low internal turnover between Directorates-General (Geys et al., 2023). In similar vein, official data from Statistics Norway indicate that in the period 2012–2017 internal mobility between ministries was circa 3% on an annual basis, while exits to the private sector were limited to circa 2% annually (https://www.ssb.no/en/statbank/list/regsys). While similar data could not be obtained at the US state level, the American State Administrators Project reports an average tenure at the state administration and agency levels of 14 and 11 years, respectively (with a maximum exceeding 50 years at both levels). As such, even in a setting where at least some positions at the top of the (state) administration are by appointment or even election, many people will still be working at the same agency or in the same (state) public administration when the next iteration of the survey is fielded.

Drop-out and repeat responders

A third and final hurdle is that at least some survey respondents must have a positive probability of answering the same survey across multiple waves. If not, it becomes de facto impossible to observe the same respondent more than once. In practice, it seems uncontroversial to assume that at least some people will not show a staunch disinclination for repeated participation. While this implies that observing these respondents repeatedly over time is feasible, the possibility that such individuals are different from their colleagues in some (un)observed—but potentially important—respects is discussed further below.

From a purely statistical perspective, we can again rely on probability theory to calculate the shares of a targeted respondent pool that are likely to answer more than once, exactly once, or never, based on the response rates in consecutive data collection efforts. To illustrate this, the response rates documented for the 1984 and 1988 survey waves of the American State Administrators Project were 38% and 48%, respectively (Yackee & Yackee, 2021). Assuming that both samples are independent draws from the exact same respondent pool,[6] this implies that 32% (i.e. 62% × 52%) of the respondent pool will not respond to any survey. In contrast, 50% (i.e. 38% × 52% + 62% × 48%) would be expected to respond once and 18% (i.e. 38% × 48%) to answer both surveys. With a universe of circa 3000 potential respondents, this equals approximately 540 individuals. Using the same calculation, response rates consistently above 60% for the Norwegian Administration Surveys (Christensen et al., 2018) would imply that 36% of targeted respondents in Norway can be expected to answer two consecutive surveys under ideal circumstances. Given a total population of roughly 3000–5000 ministerial employees, this would be in excess of one thousand respondents (i.e. 3000 × 0.36 = 1080; 5000 × 0.36 = 1800). Clearly, the potential size of a panel dataset extracted from repeated cross-sectional surveys thus will be a positive function of both survey response rates and the size of the targeted respondent pool. That is, low response rates can be compensated to some extent by a large respondent pool, and vice versa. For instance, Geys et al. (2023) report response rates in the 2008 and 2014 surveys among European Commission staff of 14% and 21%, respectively. Although this implies that only 3% (i.e. 14% × 21%) of the respondent pool would be expected to answer both surveys, this still constitutes 375 individuals given that circa 12,500 individuals were targeted.

At this point, it is important to note that it may take a certain kind of person to answer surveys, and to answer them repeatedly. Such self-selection of respondents answering more than once can be viewed as beneficial for our methodology since it increases the likelihood of observing the same respondents during consecutive survey waves. This directly increases the statistical power available in subsequent analysis of the panel dataset extracted from the repeated cross-sectional surveys. Yet, any self-selection naturally also implies that the subsample of respondents that can be observed repeatedly over time may differ from the rest of the respondent pool in a number of respects, which will generally mitigate the *external* validity of any findings. This is not necessarily problematic for the *internal* validity of any longitudinal analyses based on these data. For instance, when a (quasi-)natural experiment divides the respondent sample into treated and untreated groups in a random manner, the study can maintain high internal validity for the assessment of the (causal) influence of this treatment on the outcomes of theoretical interest (Geys et al., 2023; Hansen & Tummers, 2020; Jilke et al., 2016). Even so, it is important to keep in mind this trade-off between internal and external validity. Scholars should remain aware that observing only *a subset of* individual respondents across repeated cross-sections will always face limitations in terms of external validity and generalizability.

Comparison with related approaches

One might wonder at this point how our methodology relates to other approaches that aim to match observations to "sufficiently similar" other observations within a specific sample or population. Within the social sciences, the most common approaches to achieve this aim include propensity score matching (Rosenbaum & Rubin, 1983) and coarsened exact matching (Blackwell et al., 2009; Iacus et al., 2011, 2012). In essence, these rely on the creation of a metric reflecting the estimated "distance" in a multidimensional covariate space between two observations (usually based on the estimation of logistic regression models; Alvarez & Levin, 2021). Recent extensions of these approaches build on machine learning methods (for an overview, see Athey & Imbens, 2019) as well as Bayesian methods (e.g., Alvarez & Levin, 2021) to improve the reliability and validity of matching procedures and resulting datasets. Although one could think of our methodology as maximizing the "propensity" that two observations across two waves of the same survey are the same (or, equivalently, as minimizing the distance between them in the multidimensional covariate space), these matching estimators aim to create balance across *groups* of individuals (i.e. treatment and control) rather than at the *individual* level (as is required in our case). This is important since working with multi-person groups allows estimating a logistic regression model predicting membership of a treatment *group* conditional on covariates. Since our aim is to predict membership of a "group" represented by one unique individual, this is not a feasible option.

Our methodology also has some basic connections to similarity or scoring matrices used in, for instance, molecular biology (Johnson & Overington, 1993; Trivedi & Nagarajaram, 2020). Akin to such scoring matrices, our approach lines up a set of "features" (in our case: individuals' background characteristics) and verifies whether two observations (in our case: respondents across two survey waves) are a match or a mismatch on each "feature." Likewise akin to scoring matrices, we can thereby opt to award varying "penalties" for any mismatches observed for distinct "features" (e.g., larger penalties for a mismatch in year of birth or gender relative to a mismatch in coarsely defined education groups). Nonetheless, our methodology differs from the common usage of these scoring matrices in that our aim does not lie in the creation of an overall similarity score (achieved by adding up all mis/matches), or an overview of the dimensions along which "mutations" occur (Johnson & Overington, 1993; Trivedi & Nagarajaram, 2020). Furthermore, scoring matrices merely provide an indication for how to score (mis) matches in relation to one fixed baseline (e.g., a specific amino acid sequence). Our approach moves beyond this by automating multiple comparisons against multiple baselines (i.e. each respondent profile in one dataset against each respondent profile in a second dataset).

Finally, we should mention that our approach also takes inspiration from the vast literature analyzing pseudo-panels derived from a series of cross-sections observed over time (Deaton, 1985; Khan, 2021; Verbeek, 2008; see also note 1). Such pseudo-panels "divide the population into a number of cohorts [or 'profiles'], being groups of individuals sharing some common characteristics" (Verbeek, 2008, p. 381). A key concern in this literature is that adding more characteristics improves the accuracy of the data (in the sense of grouping together individuals with behaviors that can plausibly be considered similar across time), but comes at the cost of leading to smaller groups (Verbeek, 2008). Our methodology takes this relationship as its point of departure, since it implies that extending the number of characteristics far enough will tend to reduce group size toward one for at least some "profiles."

ETHICAL CONCERNS AND INFORMED CONSENT

As mentioned, the method presented in the previous section works best when the number of, and level of detail in, the available background characteristics increases (Table 1). From this perspective, it is important to observe that many existing surveys among public employees include very detailed background information in part due to the explicit intent to address topics including representation, inclusion, equity as well as equality of opportunity and treatment (see, e.g., the questionnaires for civil service censuses conducted in Australia, Canada, New Zealand, and the UK). Nonetheless, inclusion of many such variables may cause unease since it elevates potential ethical concerns regarding respondent privacy, anonymity, and confidentiality. Three arguments can mitigate such concerns considerably.

First, our method relies exclusively on information available within the original datasets. No new information is added or constructed beyond the observation that someone did (or did not) answer two consecutive survey waves. This, however, is not a tangible nor a visible characteristic of respondents, and thus will not help to identify a real-world individual. Moreover, although repeated responses to a survey reveal something about a respondent's length of stay in an organization, most surveys contain variables directly capturing this information via questions on respondents' years of experience (as in the American, Australian, German, and Norwegian datasets referred to previously), or their year of entry in the organization (as in the European Commission datasets referred to previously). Hence, respondent privacy and anonymity are maintained at the same level as the original surveys. The only partial exception to this is the existence of time-varying variables that can potentially be personally identifying (such as switching between specific units within the organization between survey waves). Even in this case, however, it would require further work on the part of the data user to personally identify individuals by linking the information from the survey to other external data sources. This type of activity is generally explicitly and strictly forbidden by data stewards of restricted-use datasets (see, for instance, the *Data Transfer and Use Agreement* of the American State Administrator Project at https://asap.wisc.edu/dataset, or the *Guidelines for Access to Microdata* from Statistics Norway at https://www.ssb.no/en/data-til-forskning/utlan-av-data-til-forskere).

Second, it is straightforward—and imperative from an ethical standpoint—to implement the proposed methodology using two distinct datasets: i.e. one containing *only* respondents' background information, and one containing *only* respondents' answers to substantive questions about attitudes, values, perceptions, and/or motivations. Creating the identifiers for individual respondent profiles requires only information about background characteristics. Hence, this information can be kept fully separate from respondents' answers on substantive questions to maintain anonymity and confidentiality. Variables based on, for instance, the order in which the original responses were obtained can be used to connect both datasets when needed while maintaining full respondent confidentiality (Fernandez et al., 2015, p.391). Furthermore, the final panel dataset can easily be fully anonymized to protect respondent privacy (e.g., by reducing the level of detail in background information or removing particularly revealing information altogether). Following these procedures entails that there is never a phase where individual characteristics used to match respondents across time periods can be linked with substantive answers across time periods. Consequently, recombining the system of unique identifiers generated in one dataset with the substantive data retained in a fully distinct dataset will not be able to do harm to respondents.

Third, the method can be implemented *even when* potentially identifiable information—such as, for instance, municipality of birth or precise organizational affiliation—is "disguised" prior to making the dataset available to researchers. For instance, following a common practice among data stewards providing researchers access to information from governments' administrative records (see, for instance, Statistics Norway, 2021), one can assign a random serial number to each municipality or public organization in the dataset. This intervention helps to anonymize the dataset under analysis, while researchers can continue to link respondents across multiple survey waves even when they do not know the relation between the serial numbers and municipalities/organizations. The reason is that the equivalence of information across survey waves is all that matters, *not* the information itself. Hence, observing number "22" across two survey waves is equally useful as observing "Heidelberg" or "Foreign Office" on both occasions (assuming, of course, that "22" has the same meaning in both survey waves)[7].

Even so, it obviously will be critical to develop and maintain clear regulations regarding data collection as well as the terms of their use and re-use. Respondent consent forms should therefore stress that under no circumstances will there be attempts to identify individuals from the data, and that no data will be released that can be linked to specific individuals.[8] At the same time, respondents should be made aware that their background profile might become recognizable from a combination of their responses. Such a warning is, to the best of our knowledge, extremely rare thus far. Furthermore, clear indications should be provided about the approach to, and nature of, de-identification procedures employed prior to any public release of the data. Our method can be of some benefit here, since it offers insights into essential recoding steps on the original datasets to minimize identifiability of respondents (in terms of number and detail of variables made publicly available).

Finally, while most consent forms provided to respondents up to now already specify what the data will be used for here and now, the purpose limitation principle holds that personal data "collected for specified, explicit and legitimate purposes [should not be] further processed in a manner that is incompatible with those purposes" (European Parliament, 2016, art. 5.1. (b)). Article 89(1) of this GDPR legislation, however, explicitly states that further processing for "scientific or historical research purposes (…) shall be, subject to appropriate safeguards, in accordance with this Regulation" (European Parliament, 2016, art. 89.1). In line with our recommendations above, these safeguards "may include pseudonymisation" or "can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects" (European Parliament, 2016, art. 89.1). Even though data re-use for scientific research purposes thus is not to be considered incompatible with the initial purposes of data collection, our analysis highlights that it would be beneficial to provide respondents more detailed information about the potential future re-use of the data. This suggestion is further reinforced by the fact that repeated surveys by the same public organization are often explicitly intended to engage in 'benchmarking and tracking progress over time' (as mentioned in, for instance, the objectives of the 2022/2023 Canadian Public Service Employee Survey). Hence, it would be important to specify whether and how re-use includes any recombination with other datasets—including, but not necessarily limited to, other waves of the same survey.

VALUE ADDED IN AN EMPIRICAL APPLICATION

In order to give a better idea of the potential value added of our approach, this section discusses one recently published empirical application showing how empirical inferences may be affected when moving from a "standard" cross-sectional approach to an analysis based on panel-data identified through our methodology (Murdoch et al., 2023). The original study contributes to a persistent debate–raging at least since the 1960s–on the socializing potential of international organizations by asking whether or not individuals acquire more internationalist attitudes while working in such an organization (Checkel, 2005; Haas, 1964; Sewell, 1966). It employs data from two large-scale surveys within the European Commission in 2008 ($N = 1901$) and 2014 ($N = 2209$), whereby 165 respondents could be matched across both survey waves.

Lacking longitudinal data, most previous work addressing this research question employs individuals' tenure within the organization as the central independent variable. The underlying hypothesis is that socialization effects should become reflected in a positive relationship between tenure length and (the strength of) internationalist attitudes. Empirical evidence for this positive association is at best mixed. Consequently, a widely held view developed that there exists no strong or consistent evidence of international socialization. Using a cross-sectional approach including all respondents with available tenure data in both survey waves ($N = 3446$), Murdoch et al. (2023) likewise find only a weak and statistically insignificant relation between tenure length and internationalist attitudes ($t = .92$; $p > .10$). This holds also when restricting the sample to staff members experiencing no structural organizational changes (such as mergers or divisions of their unit) between both survey waves, which arguably offers a best-case scenario for socialization effects to develop. In sharp contrast, an analysis based on panel-data identified through our methodology indicates a statistically significant strengthening of internationalist attitudes over time for individuals experiencing no structural changes relative to those who do experience such changes ($t = 2.21$; $p < .05$; full details in Murdoch et al., 2023). Hence, despite the much smaller sample size, this analysis indicates that international organizations do in fact influence attitudinal developments in individual bureaucrats. The difference between both sets of findings arises in part because cross-sectional data are unable to control for individual-level heterogeneity and potentially selective early exits linked to individuals' internationalist attitudes, which may lead to incorrect inferences (King et al., 1994; Pandey, 2017; Stritch, 2017).

CONCLUSION

Taking inspiration from scoring matrices used in molecular biology (Johnson & Overington, 1993; Trivedi & Nagarajaram, 2020) and the econometrics literature on pseudo-panels (Deaton, 1985; Verbeek, 2008), this article proposed a method to uncover individual-level panel data within repeated cross-sectional surveys of the same target population. The approach exploits that the availability of sufficiently detailed background characteristics within consecutive waves of the same survey can enable the identification of reappearing unique respondent profiles over time. While any ethical concerns regarding privacy, anonymity and confidentiality can—and should—be handled with the utmost sensitivity, the resulting panel dataset(s) offer scholars as well as practitioners vital new opportunities for longitudinal public administration research *at the individual level*. Combined with an appropriate research design, these datasets can ensure high internal validity and greater potential for credible causal inferences when testing theoretical hypotheses relative to cross-sectional analyses (Geys et al., 2023; Hansen & Tummers, 2020; Jilke et al., 2016; Murdoch et al., 2023; Stritch, 2017), even though external generalization to a broader population should be treated with due caution. Consequently, applications of our methodology to existing survey datasets can help in the further development of key concepts and *dynamic* theories in Public Administration and Public Management, including the elaboration of these theories' scope conditions as well as the verification of underlying causal mechanisms (Jensen et al., 2019; Pedersen & Stritch, 2018; Trein et al., 2021).

We acknowledge that some might remain ill at ease about our methodology irrespective of any safeguards that could be put in place. Nonetheless, it also offers practitioners, researchers, and data controllers a new tool that can provide guidance for de-identification processes of survey data when preparing Public Release datasets. As suggested by our simulations, it will in practice nearly always be unavoidable that at least some respondent profiles become recognizable from a combination of background information. Yet, data controllers can set a target or limit for how many of these cases can be deemed "acceptable," and use our methodology to assess the type of adjustments required to achieve this aim. Note that this same benefit is *not* offered by the methodological tools available in the literature on scoring matrices, pseudo-panels, and matching estimators. The main reason is that these existing methodologies are predominantly geared toward generating a diversity (or similarity) score relative to one specific baseline (e.g., membership of a treatment group in matching estimators, or a specific amino acid sequence for scoring matrices). Unlike our methodology, they cannot straightforwardly be employed to calculate the number of profiles that become recognizable depending on the number, detail and intercorrelation of the "features" included in the dataset.

ACKNOWLEDGMENTS

I am very grateful to Hussein Kassim and Sara Connolly, Falk Ebinger, Susan Webb Yackee as well as the Norwegian centre for research data (NSD) for access to relevant datasets, and greatly appreciated insightful feedback from three anonymous reviewers, Agustín Casas, Sara Connolly, Falk Ebinger, Nathan Favero, Jon Fiva, Nataliya Gerasimova, Colin Green, Georg Ibsen Lexander, Muiris MacCarthaigh, Ken Meier, Zuzana Murdoch, Hussein Kassim, Susana Peralta, Maria Petrova, Stephen Szaura, Rune Sørensen, as well as participants at two University of Bergen PhD courses (December 2021 and June 2023) and workshops/seminars at BI Norwegian Business School (Campus Bergen), NOVA Business School Lisbon and NTNU Trondheim. Kevin Huynh Tran and Henrik Nicolajsen supplied outstanding research assistance, while Alex Yeandle provided crucial support in the development of the R script. The European Commission provided financial support via the Horizon Europe 'DemoTrans' project (grant no. 101059288; Work Package leader: Zuzana Murdoch). Views and opinions expressed are those of the author only and do not necessarily reflect those of the European Commission nor the Horizon Europe program. Neither the European Commission nor the granting authority can be held responsible for them. Everlasting gratitude to Josée van Dijck.

ENDNOTES

[1] Some scholars have proposed to create so-called pseudo-panels by exploiting that individuals are embedded in social groups (e.g., departments/units or hierarchical levels) and/or share specific background characteristics (e.g., gender or race/ethnicity). By clustering individuals within such groups and calculating intra-group means, it is possible to follow these groups' development across repeated cross-sections (Deaton, 1985; Verbeek, 2008; Bertelli et al., 2015). Nonetheless, changes in intra-group means do not necessarily connect one-to-one with changes at the individual level due to, for instance, shifts in group composition across survey waves. Furthermore, while larger groups are critical to limit measurement error when calculating intra-group means, this undermines precision in statistical models by reducing the number of observations (i.e. a bias-variance trade-off).

[2] Note also that any lack of discriminatory power from having access to few background variables will not be fully compensated by having more answer categories for the available variables. In other words, five variables with seven categories offers *less* discriminatory power than seven variables with five categories (see Online Appendix Tables B.1 and B.2).

[3] Although unrelated to the functionality of the method, characteristics to be used as dependent variables in later analyses should ideally be excluded while performing the procedure to detect repeated respondents. This may risk a degree of selection on outcome variables, which could bias inferences in subsequent analyses.

[4] From this perspective, it is important to note that the approach outlined here does not violate standard policies on the use of such data. We return to this in more detail below, while discussing ethical considerations and the potential for personal identification of individuals.

[5] The Federal Employee Viewpoint Survey predominantly relies on a (stratified) random sample. Yet, since 2011 "a census was taken of employees in (…) most of the smaller independent agencies as well as 13 of the larger ones" (Fernandez et al., 2015; p.387). This increases the potential for extracting individual-level panel data within at least these units.

[6] This entails the potentially unrealistic assumption that there is no self-selection into repeated participation. We return to this below.

[7] Introducing random serial numbers does not affect the ability to control for important background variables in regression models (e.g., using indicator variables for municipalities/organizations), or cluster standard errors at, say, the level of the organization. Note also that this possibility helps address a Lucas-type critique that survey organizations, respondents and ethical committees might change their stance on survey-based research following the application of the method described here (Lucas, 1976). I am very grateful to Agustín Casas for pointing this out.

[8] Remember that our approach does not set out to personally identify specific respondents, which would always require further work by the data user. Naturally, this is ethically as well as legally unacceptable (see above).

REFERENCES

Alvarez, Michael, and Ines Levin. 2021. "Uncertain Neighbors: Bayesian Propensity Score Matching for Causal Inference." https://doi.org/10.48550/arXiv.2105.02362.

Athey, Susan, and Guido Imbens. 2019. "Machine Learning Methods that Economists Should Know about." *Annual Review of Economics* 11: 685–725.

Bertelli, Anthony, Dyana Mason, Jennifer Connolly, and David Gastwirth. 2015. "Measuring Agency Attributes with Attitudes across Time: A Method and Examples Using Large-Scale Federal Surveys." *Journal of Public Administration Research and Theory* 25: 513–44.

Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro. 2009. "Cem: Coarsened Exact Matching in Stata." *The Stata Journal* 9(4): 524–46.

Checkel, Jeffrey. 2005. "International Institutions and Socialization in Europe: Introduction and Framework." *International Organization* 59: 801–26.

Christensen, Tom, Morten Egeberg, Per Lægreid, and Jarle Trondal. 2018. *Sentralforvaltningen. Stabilitet og Endring gjennom 40 År (Central Government Administration. Stability and Change over 40 Years)*. Oslo: Universitetsforlaget.

de Geus, Christa, Alex Ingrams, Lars Tummers, and Sanjay Pandey. 2020. "Organizational Citizenship Behavior in the Public Sector: A Systematic Literature Review and Future Research Agenda." *Public Administration Review* 80(2): 259–70.

Deaton, Angus. 1985. "Panel Data from Time Series of Cross-Sections." *Journal of Econometrics* 30(1–2): 109–26.

Ebinger, Falk, Sylvia Veit, and Bastian Strobel. 2022. "Role Models in the Senior Civil Service: How Tasks Frame the Identification of Senior Bureaucrats with Active and Reactive Roles." *International Journal of Public Administration* 45(14): 991–1002.

European Parliament. 2016. "Regulation 2016/679 of the European Parliament and of the Council on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)." *Official Journal of the European Union*, L119/1 https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

Feeney, Mary, and Justin Stritch. 2019. "Family-Friendly Policies, Gender, and Work-Life Balance in the Public Sector." *Review of Public Personnel Administration* 39(3): 422–48.

Fernandez, Sergio, William Resh, Tima Moldogaziev, and Zachary Oberfield. 2015. "Assessing the Past and Promise of the Federal Employee Viewpoint Survey for Public Management Research: A Research Synthesis." *Public Administration Review* 75(3): 382–94.

Geys, Benny, Per Lægreid, Zuzana Murdoch, and Jarle Trondal (2023). "Are Civil Servants Sensitive to Terrorism? Longitudinal Evidence from the 22 July 2011 Terror Attacks in Norway." *Public Administration Review*, forthcoming.

Haas, Ernst. 1964. *Beyond the Nation State*. Stanford: Stanford University Press.

Hansen, Jesper Asring, and Lars Tummers. 2020. "A Systematic Review of Field Experiments in Public Administration." *Public Administration Review* 80(6): 921–31.

Hattke, Fabian, Rick Vogel, and Judith Znanewitz. 2018. "Satisfied with Red Tape? Leadership, Civic Duty, and Career Intention in the Military." *Public Management Review* 20(4): 563–86.

Iacus, Stefano, Gary King, and Giuseppe Porro. 2011. "Multivariate Matching Methods that Are Monotonic Imbalance Bounding." *Journal of the American Statistical Association* 106(493): 345–61.

Iacus, Stefano, Gary King, and Giuseppe Porro. 2012. "Causal Inference without Balance Checking: Coarsened Exact Matching." *Political Analysis* 20(1): 1–24.

Jensen, Ulrich Thy, Lotte Bøgh Andersen, and Christian Bøtcher Jacobsen. 2019. "Only when we Agree! How Value Congruence Moderates the Impact of Goal-Oriented Leadership on Public Service Motivation." *Public Administration Review* 79(1): 12–24.

Jilke, Sebastian, Steven Van de Walle, and Soonhee Kim. 2016. "Generating Usable Knowledge through an Experimental Approach to Public Administration." *Public Administration Review* 76(1): 69–72.

Johnson, Mark, and John Overington. 1993. "A Structural Basis for Sequence Comparisons: An Evaluation of Scoring Methodologies." *Journal of Molecular Biology* 233(4): 716–38.

Kassim, Hussein, John Peterson, Michael Bauer, Sara Connolly, Renaud Dehousse, Liesbet Hooghe, and Andrew Thompson. 2013. *The European Commission of the 21st Century*. Oxford: Oxford University Press.

Khan, Rumman. 2021. "Assessing Sampling Error in Pseudo-Panel Models." *Oxford Bulletin of Economics and Statistics* 83(3): 742–69.

King, Gary, Robert Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Lucas, Robert. 1976. "Econometric Policy Evaluation: A Critique." In *The Phillips Curve and Labor Markets*, edited by Karl Brunner and Alan Meltzer, 19–46. Amsterdam: North-Holland.

McKinney, Earl (1966). "Generalized Birthday Problem." *The American Mathematical Monthly* 73(4):385–387.

Murdoch, Zuzana, Sara Connolly, Hussein Kassim, and Benny Geys. 2021. "Legitimacy Crises and the Temporal Dynamics of Bureaucratic Representation." *Governance* 35(1): 65–82.

Murdoch, Zuzana, Sara Connolly, Hussein Kassim and Benny Geys (2023). "Employee Responses to Staff Reallocations in Public Sector Organizations: Evidence using Longitudinal Data from the European Commission." *Public Management Review*, forthcoming.

Pandey, Sanjay. 2017. "Theory and Method in Public Administration." *Review of Public Personnel Administration* 37(2): 131–8.

Pedersen, Mogens Jin, and Justin Stritch. 2018. "RNICE Model: Evaluating the Contribution of Replication Studies in Public Administration and Management Research." *Public Administration Review* 78(4): 606–12.

Ritz, Adrian, Gene Brewer, and Oliver Neumann. 2016. "Public Service Motivation: A Systematic Literature Review and Outlook." *Public Administration Review* 76(3): 414–26.

Rosenbaum, Paul, and Donald Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41–55.

Scott, Patrick, and Sanjay Pandey. 2005. "Red Tape and Public Service Motivation: Findings from a National Survey of Managers in State Health and Human Services Agencies." *Review of Public Personnel Administration* 25(2): 155–80.

Sewell, James. 1966. *Functionalism and World Politics*. Princeton: Princeton University Press.

Statistics Norway. 2021. *Access to Microdata* https://www.ssb.no/en/data-til-forskning/utlan-av-data-til-forskere.

Steijn, Bram, and Joris Van der Voet. 2019. "Relational Job Characteristics and Job Satisfaction of Public Sector Employees: When Prosocial Motivation and Red Tape Collide." *Public Administration* 97(1): 64–80.

Stritch, Justin. 2017. "Minding the Time: A Critical Look at Longitudinal Design and Data Analysis in Quantitative Public Management Research." *Review of Public Personnel Administration* 37(2): 219–44.

Trein, Philipp, Robbert Biesbroek, Thomas Bolognesi, Guillermo Cejudo, Robert Duffy, Thurid Hustedt, and Iris Meyer. 2021. "Policy Coordination and Integration: A Research Agenda." *Public Administration Review* 81(5): 973–7.

Trivedi, Rakesh, and Hampapathalu Adimurthy Nagarajaram. 2020. "Substitution Scoring Matrices for Proteins: An Overview." *Protein Science* 29(11): 2150–63.

Verbeek, Marno. 2008. "Pseudo-Panels and Repeated Cross-Sections." In *The Econometrics of Panel Data*, edited by L. Mátyás and P. Sevestre, 369–83. Heidelberg: Springer Verlag.

Vigoda-Gadot, Eran, and Itai Beeri. 2011. "Change-Oriented Organizational Citizenship Behavior in Public Administration: The Power of Leadership and the Cost of Organizational Politics." *Journal of Public Administration Research and Theory* 22(3): 573–96.

Vigoda-Gadot, Eran, and Dana Vashdi, eds. 2020. *Handbook of Research Methods in Public Administration, Management and Policy*. Cheltenham: Edward Elgar Publishing.

Wendl, Michael. 2003. "Collision Probability between Sets of Random Variables." *Statistics & Probability Letters* 64(3): 249–54.

Wynen, Jan, Bjorn Kleizen, Koen Verhoest, Per Lægreid, and Vidar Rolland. 2020. "Keeping a Watchful Eye in Times of Turmoil? How Repeated Structural Reform Leads to More Attention to Political Signals." *Public Administration* 98(3): 570–90.

Yackee, Jason, and Susan Yackee. 2021. "The American State Administrators Project: A New 50-State, 50-Year Data Resource for Scholars." *Public Administration Review* 81: 558–6.

AUTHOR BIOGRAPHY

Benny Geys is a Professor in Economics at Norwegian Business School (BI), campus Bergen. His research focuses on political behavior, (local) public policy, and public administrations. Some of his work has appeared in, e.g., *American Political Science Review*, *Economic Journal*, *Journal of the European Economic Association*, *Journal of Politics*, *Journal of Public Administration Research and Theory*, *Journal of Public Economics*, *Leadership Quarterly*, and *Public Administration Review*.
benny.geys@bi.no

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Geys, Benny. 2023. "Fancy Seeing You Here…again: Uncovering Individual-Level Panel Data in Repeated Cross-Sectional Surveys." *Public Administration Review* 1–11. https://doi.org/10.1111/puar.13693