



Examining How Conditioning on Different Wave Lengths Alters Sample Characteristics and Results in a Panel Dataset of Youth Who Have Committed Serious Offenses

Matthew C. Kijowski¹ · Theodore Wilson¹

Received: 29 January 2022 / Revised: 24 May 2022 / Accepted: 11 July 2022 /
Published online: 5 August 2022

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

Research on missing data in panel datasets has focused on attrition where respondents drop out and do not return. An equally important, but understudied, form of missing data include missed interviews where respondents contribute different number of total interviews to a panel dataset. Because individuals with crime-related characteristics miss more interviews, conditioning one's sample on different number of waves changes the composition of the sample, and likely the subsequent conclusions. Scholars must weigh the balancing act of wanting a sample with more waves of data to tease out their panel process against the concern that they will lose individuals sensitive to the process under study by requiring too many waves of data. Using the Pathways to Desistance study, a panel dataset of youth who have committed serious offenses, we compared sample characteristics across multiple conditioned samples to unpack how the characteristics of one's sample changes as more waves are required to be entered into the sample. We further demonstrate the implications of this in an applied setting by examining the relationship between residential mobility and perceptions of informal social costs. Our results indicate that the characteristics of one's panel sample are sensitive to the number of waves one conditions their sample upon. This is especially prevalent for offending where those who contribute fewer waves of data consistently report higher levels of offending. In our empirical demonstration, substantive conclusions change across conditioned samples. Our study provides unique insight into an understudied phenomenon while also providing practical advice to panel dataset scholars.

Keywords Panel data · Longitudinal data · Missed interviews · Attrition · Selection bias

✉ Matthew C. Kijowski
mkijowski@albany.edu

¹ School of Criminal Justice, University at Albany, SUNY, Albany, NY, USA

Introduction

Panel datasets have been increasingly relied upon to test criminological theories including life-course theories of crime. Panel datasets contain subjects, like individuals, neighborhoods, or states, that are repeatedly measured over time at set intervals (e.g., the same individuals are interviewed every 6 months). Panel datasets afford numerous benefits to researchers including an enhanced ability to temporally order constructs of interest, assess change and development over time among respondents in those constructs, and the added benefit of parsing out time-stable and time-varying heterogeneity among those respondents in the constructs of interest (Allison, 2009; Caruana et al., 2015; Joshi, 2016). For many criminological inquiries, these are not inconsequential benefits but rather necessary features due to the nature of the theory being tested. For example, life-course theories of crime are inherently developmental: As individuals progress from adolescence into adulthood, they become more strongly attached to agents of informal social control—such as significant others and employment—that make offending less likely because it may jeopardize their relationship with these agents (Maruna, 2001; Sampson & Laub, 1993, 2016). In order to test this theory and perspective, scholars require data on individuals that extend for several years to assess how individuals' attachments to these agents change over time and to be sure that these attachments occur before hypothesized reductions in crime.¹

Even though panel datasets have several features that make them desirable or necessary in testing crime-related theories, they also contain unique missing data patterns that may pose problems. Most criminologists are likely aware of item nonresponse where a respondent may make a mistake or decline to answer a question (e.g., income). The additional missing data pattern for panel datasets is a missed interview or multiple missed interviews (sometimes called “wave nonresponse”).² Criminologists have typically coined this as attrition or dropout. However, the conception of attrition or dropout in the singular where a respondent exits a study in full does not fully capture the nature of missed interviews in panel datasets.³ There are additional patterns of missed interviews (Lutgig, 2014; Satherley et al., 2015) where respondents contribute different number of waves which may, for example, lead one to be observed during waves 1, 3, and 5, but missing from waves 2, 4, and 6 in a six-wave panel dataset. These varying patterns of missed interviews affect the total number of interviews respondents contribute to the pooled panel dataset.

As with the well-known issue of item nonresponse, researchers must decide what to do with respondents that are missing interviews. This is often a subjective decision where researchers decide how many interviews a respondent must have contributed

¹ We are simplifying this theoretical perspective in the interest of parsimony and demonstration purposes.

² We use the terms “interview,” “wave,” “period,” “recall period,” and “follow-up interview” interchangeably throughout this paper to refer to subsequent points of data collection within a panel dataset.

³ There are instances where this characterization is accurate, such as when a respondent retracts consent or declines to continue participating in a study. This form of dropout would be a singular and total discontinuation with the study.

to be retained in their analytic sample. If this decision is not made by the researcher, computer programs applying certain panel data methods will make that decision for the researcher.⁴ Consider a panel dataset that contains six follow-up interviews after a baseline interview from which you wish to study some phenomenon of interest. What would the optimal number of waves be from which to condition the sample: at least two observed waves? at least three observed waves? observed for all six waves? This decision is complicated by a delicate balancing act scholars must consider: include as many waves of data as possible to attain a more precise depiction of the changing process weighed against the concern of excluding those who were unable to contribute the desired number of waves. These many sampling possibilities and balancing concerns are emblematic of “researcher degrees of freedom” (Gelman & Loken, 2013) where the researcher has many subjective decisions available to them, without much empirical guidance as to the consequences of these decisions. As a result, one’s ultimate sampling decision is often not stated in the manuscript whether intentionally or, more often, unintentionally.⁵ Of concern, then, is what factors determine or affect those who missed their interviews. If the primary constructs of interest *are also* related to missing an interview or multiple interviews, then this seemingly innocuous sampling decision becomes consequential.

A rich literature has shown that respondents who dropout of studies *are* different on important background, social, and psychological characteristics from respondents who remain in the sample (Fitzgerald, 2011; Hauser, 2005; Keyes et al., 2020; Thimasarn-Anwar et al., 2014). Those who attrit, or drop out, are also different in past offending experiences and other characteristics related to offending ([e.g., psychological and familial issues] Cordray & Polk, 1983; Farrington et al., 1990; Thornberry et al., 1993; Western et al., 2016). While insightful, the vast majority of criminological studies examining the effects of attrition have typically been focused upon attrition in a singular fashion where the dropout is permanent and persists for the remaining duration of the study. However, missed waves in panel datasets do not always neatly conform to this dichotomy (see, e.g., Saiepour et al., 2016; Satherley et al., 2015; Ware et al., 2006). As Lugtig (2014, p. 717) noted: “Almost all of the

⁴ For example, a fixed effects model (Allison, 2009) requires at least two waves from a respondent in order for the analysis to be conducted. The default for many computer programs is that they will include those individuals who were observed at least twice throughout the entire panel dataset. In a six-wave panel dataset, this may result in respondents being entered into the sample who were observed for any combination of two waves, including only the first and sixth waves, the first and second waves, the fifth and sixth waves, and so on. The wave combinations these individuals were observed for may not be conducive to the developmental theory under study. Other modeling strategies focused upon growth or change as one ages—such as growth curve and group-based trajectory models—often require multiple *consecutive* waves of data from respondents where those who do not contribute these required *consecutive* waves are not included. While the implications surrounding consecutive waves are beyond the scope of this study, it is conceptually related vis-à-vis the number of waves respondent contribute to a panel dataset.

⁵ Most often what is found in the literature is of the sort “we use data from a longitudinal design between the years [xxxx] and [xxxx].” Much information remains unknown, such as how many waves the researcher(s) condition their sample on and the characteristics of the individuals who are missing interviews and unable to contribute to the sample as compared to those who do provide the sufficient number of interviews to be in the sample.

respondents in our study miss one or more waves of the study. Sometimes, wave nonresponse leads to permanent dropout, but more often, respondents return to the panel survey.” As researchers require respondents to have been interviewed for more waves to be entered in their sample, there is an increased risk that the respondents comprising these extra waves of data “will not maintain the initial sample over time” (Jeličić et al., 2009, p. 1195). Those most likely to miss interviews are more likely to be included in less restrictive samples (e.g., at least two or three waves out of six total) as compared to more restrictive samples (e.g., at least five or all six waves out of six total). This is particularly salient for criminological studies because individuals with crime-related characteristics such as mental illness, justice involvement, or drug use miss more interviews (Cordray & Polk, 1983; Farrington et al., 1990; Thornberry et al., 1993; Western et al., 2016). The more waves of interviews one requires respondents to have contributed, the less likely respondents with unique crime-related characteristics are included in the sample.

Common solutions to missing data often include imputation methods, which have grown in popularity among criminologists since Allison’s (2002) exposition.⁶ These methods used to address missing data require that the researcher is aware of what is causing the data to be missing (missing at random [MAR]) or that the missing data is missing completely at random (MCAR). Imputation strategies are used with greater success to combat item nonresponse when the researcher has extensive information on the respondent from the *same* interview that is missing a few items. However, Allison (2002) notes that the reasons behind attrition in panel datasets are rarely known, and researchers then do not have the benefit of other observed items during the missed interviews to aid imputation strategies. Since these respondents report no information for the wave(s) they missed, and are likely to be different from the conditioned sample in ways related to the criminological phenomenon under study, imputation methods cannot properly restore the unique characteristics of the lost individuals. Thus, Allison (2002, p. 2) suggested that “[t]he only real good solution to the missing data problem is not to have any.” While we agree with Allison, it is not necessarily practical to researchers using panel datasets where attrition and missed interviews are commonplace and also liable to be generated by non-random processes for criminological inquiries (Western et al., 2016).

Coupling crime-related missing data processes with the researcher degrees of freedom in constructing one’s analytic sample based upon a number of observed waves, it is important that scholars understand how sample characteristics and empirical conclusions can change based upon this decision. We provide insight into these untapped issues by demonstrating how the sample and empirical results change across conditioned samples. We do this by assessing which baseline interview characteristics are associated with respondents being interviewed for differing number of waves that occur after the baseline interview. Baseline characteristics are not affected by missed interviews since all respondents reported information when

⁶ Of course, there are other methods one may use, including selection models (Heckman, 1979; see Mitchell et al., 2021, for a recent exposition on imputation and selection methods for panel data in a criminological context).

entering the study. This allows us to assess what characteristics at the baseline interview are associated with respondents reporting different number of interviews in the *future*. Furthermore, researchers using panel datasets often condition their sample on those observed for *at least* a certain number of waves (e.g., *at least* two waves; *at least* three waves), as compared to an exact number of waves (e.g., exactly two waves; exactly three waves). Unfortunately, the latter has received the most amount of scholarly attention but provide limited insight into differences across those observed for *at least* a certain number of waves that are used in empirical studies.

Afterward, we test these sampling implications by examining the effect of residential mobility on perceptions of informal social costs. We study this relationship because of its long history and continued relevance in being an important factor in life-course and development processes (Elder & Giele, 2009; Kirk, 2020), and because, as with most crime-related processes, individuals most associated with the primary variables of interest are harder to be re-interviewed. One's statistical conclusions, then, may be sensitive to the conditioned number of waves. This empirical test provides insight into the consequences associated with the balancing act of wanting more waves of data against the concern of losing those with characteristics pertinent to the relationship under study.

We use data from the Pathways to Desistance study to examine these processes. The Pathways study is a multi-year panel dataset comprised of adolescents who have committed mostly felony offenses. This dataset is useful for our purposes here, as it is a dataset with a vast array of background, psychological, social, familial, and offending variables available on all respondents regardless of prospective missed interviews. These rich data allow for a deep investigation into the factors associated with reporting differing number of waves among a sample with experiences of serious offending. This is important given the continued demand for datasets comprised of those who have offended, as these datasets are typically less scrutinized given the demand for such samples. Finally, because the study investigators devoted significant effort to prevent missed interviews (Schubert et al., 2004), differences in results between those who miss interviews and those who are observed are likely to be real differences rather than discrepancies that are confounded by the amount of effort investigators devoted to re-interview respondents.⁷

Panel Designs in Criminology

The origins of panel research in criminology can be traced back to the work of Sheldon and Eleanor Glueck.⁸ The Gluecks followed juvenile delinquents into adulthood noting that many did not desist from offending. To them, this indicated that these individuals

⁷ See Cotter et al. (2005) for the relationship between effort given to re-interview participants and selective attrition.

⁸ While the Gluecks' research is the most commonly recognized panel design in criminology, other panel designs studying delinquency were being conducted at around same time, including the Cambridge-Somerville Youth Study (McCord & McCord, 1959; Powers & Witmer, 1951) and Minnesota Multiphasic Personality Inventory Prediction Study [MMPI] (Hathaway & Monachesi, 1953).

had “criminal careers” (Glueck & Glueck, 1930, 1950, 1968). Alfred Blumstein, Jacqueline Cohen, and colleagues advanced the Gluecks’ research by bringing the concept of “criminal careers”—and the need to understand developmental trajectories—into mainstream criminology (Blumstein & Cohen, 1979; Blumstein et al., 1978, 1982, 1986).⁹ With the advent of computers able to perform complex statistical analyses, along with policy prescriptions to “selectively incapacitate” (Blumstein, 1983; Cohen, 1984) those predicted to have criminal careers at a time of exponential growth in incarceration (see National Research Council, 2014), scholars began to focus their research on panel datasets to more accurately model the processes under study.

Even with the benefits of panel datasets that were mentioned previously, critics from the past (Gottfredson & Hirschi, 1986, 1987, 1988)¹⁰ and present (Cullen et al., 2019) remind scholars of the perils of uncritically relying on these benefits, particularly in terms of funding with the number of cross-sectional studies that could be funded against the expense of a single longitudinal study. Juxtaposed with the advantages are disadvantages, including time gaps between follow-ups that are too long to render cause-effect statements, financial costs of data collection, and failure to account for the correlations of observations repeatedly measured over time (Caruana et al., 2015).

Missed Interviews and Its Consequences

A complication with panel datasets—which is also a potential disadvantage—pertains to respondent attrition. Attrition, in its truest form, is a singular occurrence where a respondent drops out of a panel dataset and does not return. Examples of this departure from panel datasets are evident in fields such as psychology and medicine where respondents withdraw their consent from further evaluation. While this is also an issue in criminology, an additional concern is that respondents in crime-related studies may be hard-to-reach and, thus, miss interviews and return at various points. This might occur if someone were to change residences and failed to update their address before the next interview or were previously incarcerated where the difficulties preceding and following incarceration make these individuals difficult to re-interview—people Pettit (2012) refers to as the “invisible men.” These respondents do not attrit or withdraw consent in the singular where they leave and do not return to the study. Rather, they are difficult to reach for additional interviews and can instead miss interviews and then be re-interviewed (Lutgig, 2014). Respondents who miss interviews contribute fewer total waves to the panel dataset than those who did not miss as many interviews.¹¹

⁹ The 1958 Birth Cohort Study (Wolfgang et al., 1972; see also Erickson, 1973) and Cambridge Study in Delinquent Development (West & Farrington, 1973) had major impacts on longitudinal research but without the heated scholarly debates that involved Blumstein and colleagues.

¹⁰ Blumstein et al. (1988) offer retorts to Gottfredson and Hirschi’s critiques.

¹¹ For example, in a six-wave panel dataset, one respondent may miss their interview at waves 1, 4, and 5 while being interviewed for waves 2, 3, and 6. Another respondent may miss their interview at waves 1, 2, 4, and 5, while being interviewed at waves 3 and 6. In this case, the first respondent contributes three total interviews to the panel dataset while the second respondent contributes only two total interviews. Of course, there are other patterns of missed interviews that affect the total number of waves respondents contribute to a panel dataset.

Scholars mostly from fields outside of criminology have noted numerous, mostly psychological, variables that influence attrition that also have been found to be correlated with offending in the criminological literature.¹² More recently, increased focus has been placed on the factors that determine missed waves. These include patterns of missed interviews across the entire panel dataset, and more specific typologies of respondents, such as those who respond to all waves, those who respond to early waves but then miss later waves, those who miss early waves but return for later waves, and those who do not respond to any waves (see Lugtig, 2014; Saiepour et al., 2016; Satherley et al., 2015; Ware et al., 2006).

As important as these studies are to our understanding of attrition and missed interviews, there are additional factors that deserve attention for criminologists studying hard-to-reach individuals with past offending experiences. Western et al., (2016, p. 5484; see also Western, 2018) went into detail about these issues for a panel study comprised of formerly incarcerated adults:

the risk of survey nonresponse were closely linked to the social risks and vulnerabilities of scientific interest. Thus, the most likely nonrespondents had histories of drug addiction and mental illness and were also more likely to be arrested and incarcerated after prison release. Because the probability of non-interview is related to extreme social disadvantages, nonresponse is strongly nonignorable.

As Western et al. (2016) noted, those with past offending experiences, and those with past incarceration experiences, must contend with additional factors that are intertwined with social disadvantage. These high-risk populations experience hindered employment opportunities and job discrimination (Pager, 2007; Sugie et al., 2020), residential instability (Kijowski & Wilson, 2021; Remster, 2019; Western, 2018), changing phone numbers (Western et al., 2016), poverty (Pettit, 2012), fractured social support systems (Western, 2018), childcare duties (Western, 2018), and re-offending that may then lead to re-incarceration, among other factors. These experiences interact with one another resulting in social insecurity and a reduced propensity to respond to surveys (Western et al., 2016).

¹² These variables include low IQ (Beaver, 2013), being a male (Cunradi et al., 2005), being an African American (Siddiqui et al., 1996), having psychopathological issues (Badawi et al., 1999), poor education (de Graaf et al., 2000), living in an urban setting (de Graaf et al., 2000), illicit substance use (Delfabbro et al., 2017), cigarette use (Cunradi et al., 2005), having four or more psychological disorders (Eaton et al., 1992), heavy alcohol consumption (Thygesen et al., 2008), childhood aggression (Huesmann et al., 1984), low socioeconomic status (Launes et al., 2014), intergenerational socioeconomic status (Schoeni & Wiemers, 2015), juvenile justice system involvement (Thimasarn-Anwar et al., 2014), non-intact family (Watkins & Melde, 2007), low self-restraint according to the Weinberger Adjustment Inventory (WAI; Weinberger et al., 1990), property victimization (Berzofsky & Carrilo-Garcia, 2018), being punished in school (Boys et al., 2003), and residential mobility (Fitzgerald et al., 1998; Young et al., 2006) among others still. See also Olver et al. (2011) for a meta-analysis on offender treatment attrition in psychology where numerous variables that are correlated with offending, and offending itself, lead to higher rates of attrition.

Given the additional social issues associated with high-risk individuals, researchers have been actively engaged in techniques that increase the likelihood of re-interviewing the hardest to reach individuals (see Cotter et al., 2002, 2005; Schubert et al., 2004). This has become an explicit focus in recent criminological studies with a growing literature on practical methods that assist in re-interviewing high-risk individuals—such as those with past offending experiences and past incarceration experiences (Clark et al., 2020; Eidson et al., 2017; Fahmy et al., 2019; Western, 2018; Western et al., 2016, 2017). For example, researchers working on the Boston Reentry Study were able to retain 91% of their sample of 122 respondents through four strategies the researchers employed: incentives for interviews, persistent phone calls and mailed letters, secondary contacts (such as friends and family), and partnerships with the justice agencies and local community groups (Western, 2018, chapter 2; see also Fahmy et al., 2019 for detail on retention techniques employed for the LoneStar Project). Researchers' re-interviewing techniques matter so much so that Clark et al., (2020, p.22) suggested that "[...] what researchers do in terms of the operational groundwork (i.e., reminders, concerted retention, etc.) matters far more than the characteristics of who they study."

A remaining concern in criminological studies is that even with respondent retention efforts, respondent characteristics will impact the number of waves they participate in. This is a remaining concern even when the researchers collecting the data go to great lengths to follow-up on the respondents as it is extremely difficult, if not impossible, to follow-up on each respondent across all waves of the data collection. The hardest to reach individuals with heightened levels of the characteristics associated with one's research may still miss more interviews. Just as the above-mentioned psychological, sociological, and demographic variables may lead to attrition in the singular, it may also cause one to miss more interviews than respondents without those characteristics (see Eisner et al., 2019; Lugtig, 2014; Saiepour, et al., 2016; Satherley et al., 2015; Ware et al., 2006).

The dearth of research in criminology on how individuals with certain characteristics contribute different total number of waves to a panel dataset may have led to a belief that the number of waves one conditions their sample on is an innocuous decision vis-à-vis the characteristic composition of the sample as compared to other possible conditioned samples.¹³ Although this is an understandable assumption without scholarship to the contrary, we believe this to be an important issue criminologists should consider similar to when Jeličić et al., (2009, p. 1195) warned psychologists that convention up to that point had been to "assume that missing data are a natural phenomenon that needs no attention since it is ubiquitous. Nothing could be further from the truth." These seemingly innocuous decisions may lead one to sample on

¹³ While we did not survey researchers in order to make this statement, published articles tell this story. It is rare to find papers that state the number of waves the sample was conditioned on, let alone how representative that sample is in relation to the baseline sample characteristics, or how these characteristics change across different conditioned samples.

those who have different characteristics while also creating issues with replicability. In the spirit of replicability, other scholars employ their own researcher degrees of freedom and condition their panel sample on a number of waves that they believe to be the most optimal where findings may be sensitive to the characteristics of the sample across waves. Gelman and Loken (2013, p.1 [emphasis added]) may have best described this issue with regard to researcher degrees of freedom: “A dataset can be analyzed in so many different ways (with the choices being not just what statistical test to perform *but also decisions on what data to exclude or [include].*” These decisions involve the various waves within a panel dataset one chooses to condition their sample upon. Because these are largely subjective decisions liable to influence one’s conclusions (see Gelman & Loken, 2013; Simmons et al., 2011; Steegen et al., 2016), a deeper understanding of this unstudied process is needed so panel dataset scholars are aware of how this seemingly innocuous decision can have consequences for their substantive conclusions.

Residential Mobility and Perceptions of Informal Social Costs

To demonstrate how the above-mentioned processes can impact one’s results, we examine the relationship between one’s residential mobility and their subsequent perceptions of informal social costs. The relationship between moving and reduced perceptions of informal costs has a long scholarly history in explaining crime (Shaw & McKay, 1942; see also Sampson, 1991; Warner & Rountree, 1997), while also continuing to receive empirical attention (Kijowski & Wilson, 2021; Vogel et al., 2017). The primary argument arising from this discourse is that moving severs ties to social others who may impose informal punishment on the individual should they offend. Moving introduces new social others into one’s life that do not carry the same informal punishment because the mover has not had enough time to bond with them. Residential mobility has a unique influence on one’s offending when studying life-course and developmental processes (Kirk, 2020; Widdowson & Siennick, 2021; see also Horney et al., 1995).

It is noteworthy that those who move, and move more often, are also more likely to miss interviews (Fitzgerald et al., 1998; Satherley et al., 2015; Western et al., 2016; Young et al., 2006). This suggests that as we require more interviews from respondents to be entered into our sample, the sample will become less residentially mobile while also having higher levels of perceived informal social costs. Here, the above-mentioned balancing act of wanting more waves of data to tease out the effect more precisely over time against the concern of losing individuals is salient. Over time and over repeated interviews, particularly as one ages, scholars may wish to examine how the effect of moving on perceptions of informal costs change. For example, the proposed negative effect may diminish from adolescence into adulthood as residential changes are undertaken in order to solidify a more stable life, such as moving for work, education, or significant others. As compared to moving during adolescence, these types of moves likely have a different effect on perceptions of informal costs requiring extended waves to tease out the changing effects. The desire to examine this developmental process

is balanced against the concern that by requiring more waves to be in the sample, one loses those who are both more mobile and have lower levels of informal costs. This balancing act permeates criminology as many commonly studied relationships are focused upon offenders with crime-related characteristics—such as those mentioned in footnote 12 and the paragraph following footnote 12—who may be especially sensitive to missed interviews. An appraisal centered on these issues can result in more robust conclusions and increased clarity for scholars wishing to apply and build upon others' findings (Gelman & Loken, 2013; Simmons et al., 2011).

The Current Study

In this study, we address an issue that permeates all panel data analyses: missed interviews and the resulting variability in the number of waves respondents contribute to the panel dataset. All scholars using panel datasets must condition their sample on being observed for a certain number of waves. This is mostly a subjective decision involving the balancing act of wanting to include as many waves as possible against the concern that one will lose individuals with unique characteristics who were unable to contribute the desired number of waves. Of concern in a criminological context is that those unable to contribute the desired waves offend at higher rates while also having crime-related characteristics. Should one condition their sample on those observed for an extended number of waves, the sample may be different from other samples that do not require as many waves and thus affect one's substantive conclusions.

While Jeličić et al. (2009) raised the alarm about these potential pitfalls to psychologists, they did so without explicitly assessing how sample characteristics and subsequent results change across conditioned samples. We build on their work by demonstrating in a criminological context how these sampling decisions alter the characteristics of the sample and substantive conclusions. We use the Pathways to Desistance study for two primary reasons. First, the dataset contains a rich trove of demographic, personal, psychological, social, and behavioral variables that were measured of each respondent that is not often measured in other panel datasets of individuals with past offending experiences. Second, the researchers went to great lengths to re-interview as many respondents as possible (Schubert et al., 2004), which helps account for the differences in sample representativeness that may arise due to the researchers re-interviewing methods (Clark et al., 2020; Fahmy et al., 2019; Western, 2018). We compare the baseline values across respondents in multiple conditioned waves in each of the 120 variables that met our sampling criteria. We then examine the relationship between residential mobility and perceived informal social costs. This relationship has important implications in studying crime and also life-course and developmental processes, while and also tapping into the concern that individuals associated with the primary variables of interest are also liable to miss interviews.

Data

Sample

We used data from the Pathways to Desistance study which is a panel study of youth who have previously offended (Mulvey, 2012). The 1354 youth who originally entered this study were found guilty of committing serious (mostly felony) offenses in Maricopa County (Phoenix), AZ, or Philadelphia County, PA. Youth entered the study between the ages of 14 and 17. The sample was mostly non-white (80%) and male (86%), while the average age of entry into the study was 16. After initial entry into the study between 2000 and 2003, youth were interviewed every 6 months for the first 3 years and every 12 months for the 4 years thereafter (see Schubert et al., 2004 for more information). We used the first 3 years (six waves) of data because of recall biases that arise in the data when respondents are first asked to recall the events of the past 6 months and are then asked to recall the events of the past 12 months (see, e.g., Thornberry & Krohn, 2003).

The Pathways study has many ways by which missing data may arise. For each of the six waves one could have been interviewed for, respondents were either completely observed, partially observed, or not observed. Those who were completely observed contributed full information on all variables for the respective wave, while those who were partially observed did not contribute full information on at least one measure. Individuals who were not observed for a wave were not interviewed in any respect and contributed no information for that wave. Our primary interest lay in those who were completely or partially observed at each wave. In the interest of thoroughness, though, we also analyzed those who were categorized as being completely observed only, treating those who were partially observed and not observed as missing (available in section D of the online supporting material). These restrictions for our primary analyses were implemented to remain as true as possible to the concept of missed interviews wherein respondents were either observed to some extent at a given wave or were not observed at all and did not report information.

We then created conditioned samples to compare across those who were observed for different number of waves. Our first sample comparison is for those who were observed for zero or one wave ($n=37$) compared to those who were observed for at least two waves ($n=1317$). We conducted this initial comparison in order to identify differences between those who are unable to contribute at all to a panel sample against those who do contribute in some form. This is a useful starting point because any differences between these samples would signal differences in sample characteristics simply from deciding to use panel methods.

Following this initial comparison, we then created samples that most closely reflect the different sampling options scholars have available to them when selecting their panel sample. Typically, researchers choose their samples based upon respondents having contributed at least a certain number of waves (e.g., at least two waves). We compared samples that incrementally increased the required number of waves respondents must have contributed to be included in the sample.

This will demonstrate what types of individuals are being sampled upon as one requires more waves of information from respondents as compared to those who are not included in the sample because they did not contribute the necessary number of waves. The first sample comparison is comprised of those who contributed only two waves ($n=19$) against those who contributed at least three waves ($n=1298$). Here, we are comparing those who contributed the minimum number of waves to be included in a panel sample against those who contributed three or more waves. Next, we increased our wave requirement by one additional wave in comparing those who contributed two or three waves ($n=52$) against those who contributed at least four waves ($n=1265$). We similarly make the wave requirements more restrictive in our final two comparisons of those who contributed between two and four waves ($n=199$) against those contributing five or six waves ($n=1198$), and those who contribute between two and five waves ($n=278$) compared to those who contribute all six waves ($n=1039$).¹⁴ This procedure compares those who are included based on one's sampling requirement against those who *could have been* included, but were not due to the same requirement.

Measures

At the baseline interview, the publicly available dataset contained variables covering six primary domains: background characteristics, individual functioning, psychosocial development and attitudes, family context, personal relationships, and community context (see Pathways to Desistance n.d.). Beyond general item missingness (e.g., refuse), there were item-specific skip patterns that excluded respondents if they did not meet certain criteria, such as asking one about their relationships with teachers in a community school where the skip pattern excluded those not in a community school. However, some skip patterns allowed us to substitute answers so as to still be able to use the variable for analysis. We did this only when the skip pattern can be quantified with confidence, such as recoding to "0" a question asking for a count of one's four closest friends who were arrested wherein the skip pattern was for those who stated they had zero friends. We excluded variables that had unfillable skip patterns. This assures us that any differences we find are missing wave-related and not confounded by item-specific skip patterns.¹⁵

¹⁴ Note that 77% of the original sample were followed up for all six waves, 88% were followed up for at least five waves, 93% were followed up for at least four waves, 96% were followed up for at least three waves, and 97% were followed up for at least two waves. This is a testament to the efforts of Schubert et al. (2004) to limit missed interviews as much as possible. Their re-interview rate is especially commendable given the known difficulty of re-interviewing high-risk individuals (Western, 2018). This gives us more confidence that any differences we do find are unlikely to be sampling artifacts confounded by researchers' efforts to follow up on respondents.

¹⁵ We did not investigate variables with item missingness where respondents were observed for the wave but did not provide full information for specific variables—which entailed refusing, skipping, or answering "don't know." Examining item-specific patterns in addition to person-specific missed waves may confound any differences we find as it would be difficult to disentangle if differences were a result of missed interviews, item nonresponse, or both.

After making these a priori decisions regarding what variables to include, we noticed that measures of offending did not meet our criteria. However, for many of the past offending measures, but not all of them, the same three respondents had item missingness indicators of “missing data,” “refuse,” or “don’t know.” Thus, we analyzed past offending measures with these three individuals removed. The sample sizes for those observed zero and one time, and two times only remain the same, while the sample sizes for those observed at least twice, at least three times, and so forth are now comprised of three fewer respondents. We computed estimates of offending in separate tables and figures to more cleanly bifurcate when these three respondents are removed from the analyses. Given our focus on criminological inquiry and the minimal loss of respondents for these highly relevant offending variables, we found this decision to be a justifiable departure from an otherwise strict sampling criterion. These restrictions resulted in our analyses being comprised of 120 variables at the baseline interview. See section A of the online supporting material for details on each variable used for this study, where we also describe skip pattern recoding for relevant variables.¹⁶

Empirical Demonstration

The data for our empirical demonstration came from the publicly available Pathways to Desistance study, while the residential mobility information came from the restricted-access monthly calendar data that we obtained and couple with the publicly available data. Rather than the data coming from only the baseline interview that comprises our first set of analyses, the data used for our empirical demonstration are measured at each 6-month recall period between the first and final interviews. We created multiple conditioned samples comprised of those who were observed for at least two waves ($n=1315$), those observed for at least three waves ($n=1296$), those observed for at least four waves ($n=1248$), those observed for at least five waves ($n=1150$), and those observed for all six waves ($n=885$).

Dependent Variable

The dependent variable is the perceived social costs that one believes would arise from various social others should they offend and be arrested by the police. The measures came from Nagin and Paternoster (1994) and comprised six items asking each respondent if the police were to catch them breaking the law, how likely is it that he or she would lose respect from neighbors and adults, lose respect from family members, be suspended from school, lose respect from close friends, lose respect from a girlfriend or boyfriend, or find it harder to get a job. Response options ranged from 1 (very unlikely) to 5 (very likely). Averaged across all waves, each of these items loaded onto the respective wave social costs scale for an acceptable reliability of 0.77.

¹⁶ More detailed information than that provided here can be obtained at <https://www.icpsr.umich.edu/web/NAHDAP/studies/29961/datadocumentation>

Independent Variables

We analyze two primary independent variables in this study. The first is residential mobility which was obtained from the restricted-access Life Event Calendar. The measure captures the number of residences each respondent lived at for each month over the past 6 months. Every month, respondents reported what their primary residence was. A different primary residence from the month prior was considered a residential change. This monthly level data was aggregated into the 6-month period as the number of residential moves during the past 6 months.

The second independent variable is one's reported offending for the past 6 months. The measure captured whether or not each respondent committed the following crimes in the past 6 months: destroyed/damaged property, arson, burglary, shoplifting, bought/received/sold stolen property, used checks/credit cards illegally, auto theft, sold marijuana, sold other illegal drugs, carjacking, drove under the influence, paid for sex, forced sex, shot someone, shot at someone, robbery with a weapon, robbery without a weapon, beaten someone up, been in a fight, fought as part of a gang, carried a gun, and murder. We obtained the sexual assault and murder measures from the restricted-access data. Answers for each offense were either a "yes" or a "no" and resulted in a proportion between 0 and 1 as to the number of different offenses they committed. For more intuitive interpretations, we transformed the proportion into a count variable of the variety of offenses committed in the past 6 months. As compared to a frequency count of all of the crimes committed, variety scores are less influenced by high-offending outliers (Sweeten, 2012).

Control Variables

To assure that the relationship between mobility and perceived social costs is due to residential moves and not other forms of transiency, we control for forms of mobility that are not residential changes. First, we account for street time so that those who spent more time outside of an institution do not unduly affect our estimates. Second, we include a variable for the days one ran away, lived on the streets, or were at various non-residential places. Finally, we account for living in public housing in order to help account for the effect of living in a disadvantaged area. We also account for other variables that may confound the relationship under study. Witnessing violence is a count variable of the number of times one was exposed to violence in the past 6 months. Personal victimization is a count of the number of times one was victimized in the past 6 months. A binary variable captures whether one was employed or not. Gang membership was measured by a binary variable as well. Finally, anti-social peer influence was measured through 19 items that captured the extent to which one has deviant peer influences.

Across each wave, there are patterns of missingness due to missed interviews, refusal, skip patterns, and reporting "don't know." They were all treated as missing and excluded from the samples. Although this confounds missed interviews with item nonresponse, this is the most applicable in modeling panel dataset phenomenon.

Analyses

Our first set of analyses are mean difference tests comparing baseline values across the five sample comparisons for all 120 variables. Depending on the distribution of the variable, we employed either a t -test, a z -test of proportions, or a non-parametric test. With such a large number of statistical tests, we increase the likelihood of finding statistically significant differences due to chance. To ameliorate this concern, we employ three p -value corrections for multiple comparisons that range from most to least conservative, respectively. We first use the Bonferroni correction for all 600 comparisons ($p < 0.00008333$). As this threshold is very conservative and inflates the likelihood of a type II error (Reinhart, 2015), we also used the Bonferroni correction accounting for 120 comparisons ($p < 0.00041667$) for the number of variables in our study. Finally, we also use Winship and Zhuo's (2018) recommendation of $p < 0.005$. Beyond these indicators, we directly report the raw p -value for all tests for purposes of clarity and for readers who may be interested in specific comparisons. Because statistical (in)significance does not necessarily equate to a meaningful difference (no difference), we report effect sizes for all comparisons. We then conduct analyses on the average effect size within group comparisons to test for group-level effect size differences among all of the variables.

For our empirical demonstration, we use the fixed effects model (Allison, 2009) with clustered standard errors. Clustered standard errors account for dependencies from querying the same respondents over time. The fixed effects model is best employed when there is significant covariation between the time-stable factors and independent variables. A Hausman test (1978; see also Wooldridge, 2010) confirmed that there is significant covariation, making the use of fixed effects over random effects appropriate. Fixed effects completely removes between-individual variation, relying only on within-individual variation. This model removes any covariation there is between time-invariant heterogeneity and the time-varying variables. This is done through a de-meaning computation where, for each variable for each individual, the average across reported waves is taken and then the deviations from that mean at each wave are computed. The average of the mean deviations comprises the effect across all reported waves. This computation is able to remove time-stable influences because the mean of a constant is the value of that constant, and subtracting a constant from a constant is zero and is thus removed from the analysis. Only time-varying factors that affect the independent variables and dependent variable at the same time can confound the relationships under study.¹⁷

These two sets of analyses provide distinct and unique insights. The first set of analyses, employing mean difference tests, is entirely *prospective*. All tests are

¹⁷ Even though the Hausman test (1978) indicated that random effects models were biased because time-stable factors confounded the relationships, we estimated random effects models to assess how sensitive results are across conditioned samples in an estimator that is mis-specified. The model predicting social costs from mobility is volatile across conditioned samples with the random effects model. Different variables, including mobility, jump in and out of significance across each of the different sample specifications. This demonstrates the importance of proper specification, especially when the samples differ on important characteristics.

conducted for the baseline measures between those who *prospectively* contribute different number of waves. This allows us to assess what distinguishes individuals' *future* number of reported waves *solely based upon their measures at the baseline interview*. The second set of analyses, the empirical demonstration, mixes these baseline sampling differences with time-varying components. This occurs because there are the baseline sampling differences in addition to difference across individuals impacting the relationship under study across waves. Therefore, the first set of analyses provides the cleanest depiction of differences between those contributing different waves *before* they even contributed those waves, while the second set of analyses mixes this with time-varying components that provides the most applicable insight for panel dataset scholars.

Results

We report only significant comparisons among the non-offending variables in Table 1. We display only the comparisons that reached statistical significance in the main text for brevity. However, all non-significant comparisons are available in section B of the online supporting material. In our tables, the direction of the effect is placed on the effect size by assigning a negative sign when the mean is larger for those in the group representing fewer waves. For some comparisons, all members within a group reported a value of "0" for certain variables, resulting in a mean of "0" and "0" standard deviation. In these instances, effect sizes are not computable and we use "N/A" when this occurs.

Three comparisons are statistically significant when using the Bonferroni-corrected threshold for statistical significance that is adjusted for all 600 comparisons. Those interviewed between two and five waves were more likely to be interviewed in a facility rather than in the community at baseline as compared to those observed for all six waves ($d=0.312$), and they were also less likely to have been in school prior to being detained at baseline as compared to those observed for all six waves ($d=0.277$). Those observed between two and five waves were older at baseline than those observed for all six waves ($d=-0.295$).

Only one comparison reached statistical significance under the 120-comparison Bonferroni-adjusted alpha. Those interviewed between two and four waves were older at baseline as compared to those observed for five or six waves ($d=-0.372$).

Finally, seven comparisons were statistically significant under the Winship and Zhuo (2018) threshold for statistical significance of 0.005. While one of these p -values was very small (e.g., $5.9e^{-4}$), others were not (e.g., 0.003). We suggest readers use caution for some of the larger p -values as they may still be a product of multiple comparisons. Those who contribute zero or one interview—and thus are not able to be in a panel sample—are more likely to be from Philadelphia as compared to Phoenix ($d=-0.495$) and are older at baseline ($d=-0.522$). Those who contributed at least two waves as compared to those contributing zero or one wave have more people living in their household ($d=0.537$), are more likely to have their biological mother living in the household ($d=0.519$), and have more biological parents who are still living ($d=0.528$), at the baseline interview. Those who were

interviewed for five or six waves as compared to being interviewed between two and four waves were more likely to have been in school before being detained at baseline ($d=0.302$). Those observed between two and five waves reported more petitions to court prior to the baseline interview as compared to those who contributed all six waves ($d=-0.233$). The findings for each variable in Table 1 operate in the same direction regardless of whether they are statistically significant.

In Table 2, we conduct similar comparisons for only the offending variables that reached our statistical significance thresholds. The variables that did not reach statistical significance can be found in section C of the online supporting material. While no comparison met our threshold for the 600-comparison Bonferroni-corrected alpha, one comparison did for our 120-comparison Bonferroni-corrected alpha. Those observed between two and five waves as compared to all six waves were more likely to have sold drugs other than marijuana prior to the baseline interview ($d=-0.245$). Two comparisons met the Winship and Zhuo threshold. Those observed between two and five waves as compared to those observed for all six waves reported more of both variety offenses committed more than 6 months prior to baseline ($d=-0.207$) and variety offenses committed more than 6 months prior to baseline that were not drug involved ($d=-0.195$). Each comparison in this table indicates that those who were interviewed for fewer waves have committed more offenses at baseline, regardless of statistical significance. Similarly, in section C of the online supporting material, the majority of comparisons operate in the same direction where those observed for fewer waves report more past offending. Contributing fewer waves seems to be correlated with increased levels of past offending.¹⁸

As we noted earlier, a statistically (in)significant p -value does not necessarily mean a (un)meaningful difference. This is seen by the fact that our largest effect size (-0.545) had a p -value of 0.018 and did not fall below any of our corrected p -values. Using Cohen's (1988) guidelines for effect sizes for the social sciences, in Table 1 and section B of the online supporting material which comprise all non-offending variables, 165 effect sizes fall between small (0.10) and medium (0.30), 43 are between medium (0.30) and large (0.50), and 7 are greater than 0.50. The effect sizes falling within these categories in Table 2 and section C of the online supporting material for the offending variables were 43, 13, and 0, respectively. There are meaningful differences between groups that are not captured by statistical significance. To examine average effect size differences between samples, we averaged the effect size of all comparisons within each sample comparison to create a sample comparison average effect size. We use 95% confidence intervals to assess

¹⁸ Section D of the online supporting information displays results for those who were completely observed providing similar conclusions. There are some differences, though. Petitions prior to baseline, variety of offenses committed beyond 6 months ago that were not drug crimes, and variety of offenses beyond 6 months ago do not reach any of our significance thresholds for the same group comparisons as in Tables 1 and 2 (although for both offending measures those observed for zero or one wave as compared to two or more offended at a higher rate rather than those observed between two and five as compared to all six). However, those observed at least twice had more adults at home as compared to those observed for zero or one wave, and those observed between two and five times ran away from home more often than those observed for all six waves.

Table 1 Comparing mean effect sizes and *p*-values for all variables except offending across waves of observations. Significant comparisons only

Variables	Wave comparisons				
	0/1 vs. 2+	2 vs. 3+	2/3 vs. 4+	2-4 vs. 5+	2-5 vs. 6
Interview (community vs. facility)					
Effect size	.330	.138	.217	.188	.312**
<i>p</i> -value	.048	.551	.124	.050	4.5e ⁻⁶
Site (Philadelphia vs. Phoenix)					
Effect size	-.495†	-.460	-.139	-.061	-.175
<i>p</i> -value	.003	.047	.326	.528	.009
Age					
Effect size	-.522†	-.350	-.256	-.372*	-.295**
<i>p</i> -value	.002	.130	.071	1.2e ⁻⁴	1.3e ⁻⁵
Number living in house					
Effect size	.537†	.300	.190	.145	.074
<i>p</i> -value	.001	.195	.179	.131	.272
Bio. mother in house					
Effect size	.519†	.012	.072	.154	.096
<i>p</i> -value	.002	.960	.610	.110	.154
Number of Bio. parents living					
Effect size	.528†	.333	.061	.165	.131
<i>p</i> -value	.002	.149	.665	.087	.053
In school before detained					
Effect size	.158	.199	.332	.302†	.277**
<i>p</i> -value	.343	.388	.019	.002	4.5e ⁻⁵
Petitions before baseline interview					
Effect size	-.341	-.107	-.135	-.217	-.233†
<i>p</i> -value	.041	.645	.339	.024	5.9e ⁻⁴

Sample sizes: 0/1 = 37; 2+ = 1317; 2 = 19; 3+ = 1298; 2/3 = 52; 4+ = 1265; 2-4 = 119; 5+ = 1198; 2-5 = 278; 6 = 1039

†Winship and Zhuo (2018) suggested $p < .005$

*Significant at the $p < .00041667$ ($p < 4.1667e^{-4}$) with 120 Bonferroni corrections

**Significant at the $p < .00008333$ ($p < 8.333e^{-5}$) with 600 Bonferroni corrections

if the average effect sizes were different from zero within each sample comparison. This type of analyses goes beyond direct comparisons of individual variables and instead examines the larger trends.

Figure 1a contains the graph for all non-offending variables (the effect sizes from Table 1 and section B of the online supporting material). None of the five comparisons has a mean effect size that does not cross zero, although some are close. Moreover, there is not a clear direction that can indicate whether those observed for fewer number of waves have higher or lower averages as compared to those observed for more waves. Figure 1b contains the offending variables only (effect

Table 2 Comparing mean effect size and *p*-values for offending variables only across waves of observations. Significant comparisons only

Variables	Wave comparisons				
	0/1 vs. 2+	2 vs. 3+	2/3 vs. 4+	2-4 vs. 5+	2-5 vs. 6
Variety of offenses beyond 6 mo. ago					
Effect size	-.443	-.324	-.180	-.153	-.207†
<i>p</i> -value	.008	.162	.203	.113	.002
Variety of offenses committed beyond 6 mo. ago no drug crimes					
Effect size	-.432	-.319	-.136	-.158	-.195†
<i>p</i> -value	.010	.168	.338	.010	.004
Sold other drugs ever					
Effect size	-.216	-.413	-.172	-.107	-.245*
<i>p</i> -value	.194	.074	.224	.265	2.9e ⁻⁴

Sample sizes: 0/1 = 37; 2+ = 1314; 2 = 19; 3+ = 1295; 2/3 = 52; 4+ = 1262; 2-4 = 119; 5+ = 1195; 2-5 = 278; 6 = 1036

†Winship and Zhuo (2018) suggested $p < .005$

*Significant at the $p < .00041667$ ($p < 4.1667e^{-4}$) with 120 Bonferroni corrections

**Significant at the $p < .00008333$ ($p < 8.333e^{-5}$) with 600 Bonferroni corrections

sizes from Table 2 and section C of the online supporting material). Contrary to the non-offending variables, four of the five offending comparisons do not cross the zero threshold. Those observed for two waves, two and three waves, between two and four waves, and between two and five waves each have a higher offending average effect size than their higher wave counterparts. Each of the five comparisons has a clear directional effect where those observed for fewer waves have higher average effect sizes. Those more likely to miss waves offend at higher rates and are not included in samples that require respondents who have contributed extended waves of information.¹⁹

Next, we computed the absolute value of the average effect sizes. This was done to assess the effect size that is unaffected by the sign assigned to the groups. This allows us to capture the average effect size that is not weakened by comparing samples of differing signs. Figure 2a displays the average absolute value effect size for all non-offending variables. There is a clear linear pattern of decreasing average effect size from excluding those who miss more interviews to then including them. This suggests that excluding those who miss more interviews has a detrimental effect on one's sample representativeness. The absolute value average effect sizes are largest when comparing those observed zero or one time against those observed at least twice, and when comparing those observed twice against those observed at least

¹⁹ Similar findings are found in tables S-1a and S-1b of the online material for those only completely observed. However, for the non-offending variables, those observed between two and five waves as compared to six are in the same direction but have an interval that does not cross zero, while for the offending variables, the same comparisons do not cross zero except for those observed between two and four waves against those observed for five or more.

three times. Note that when focusing only on those able to contribute to panel samples (excluding the zero and one wave comparison), none of the confidence intervals overlaps with the sample comparison of two waves against three or more waves. Moreover, the comparison of two and three waves against four or more does not overlap with those observed for between two and five waves against those observed for all six waves. Figure 2b contains the graph for the absolute average effect size for the offending variables only. This figure demonstrates that excluding those who contribute fewer waves of data unduly impacts the offending nature of one's sample. Again, when considering only samples that contribute to panel samples, none of the confidence intervals overlaps with the comparisons of those observed twice against three or more times.²⁰

Finally, we combined the effect sizes for the non-offending variables and the offending variables in order to have a global assessment of the sample comparison mean effect size and absolute average mean effect sizes. Figure 3a displays these group comparisons. Two of the five effect sizes have intervals that do not cross zero, while one barely crosses zero. Specifically, those observed for two waves as compared to three or more waves and those observed between two and five waves as compared to those observed for all six waves each reported a higher average on the variables under study as compared to their counterpart sample. Figure 3b displays the absolute value mean effect sizes. There is a decreasing linear effect as those who contribute fewer waves become included. For panel samples only (excluding those observed zero or one times), the confidence intervals do not overlap with those observed twice as compared to three times or more. The comparison of those observed two and three times against those observed for four or more waves does not overlap with the comparison of those observed between two and five times against all six times—and barely overlaps with the comparison of those observed between two and four times against five or more times. Conditioning across waves affects the criminality and characteristic make-up of the sample.^{21,22}

For our empirical demonstration, we begin by taking the means of each variable across the conditioned samples in Table 3.²³ Those who contribute fewer waves have lower levels of perceived social costs, while also moving more often. As we

²⁰ Similar results are found in tables S-2a and S-2b examining the completely observed only samples.

²¹ Tables S-3a and S-3b are similar except that comparing two waves against three or more in table S-3a crosses zero.

²² In section E of the online supporting material, we compare across each group—conducting separate analyses for both complete and partial and complete only—those who contributed exactly zero or one waves, exactly two or three waves, exactly three waves, exactly four waves, exactly five waves, exactly six waves, and less than six (including zero and one) compared to all six. Although this is less applicable to panel dataset scholars who use those observed for at least a certain number of waves, it provides a more detailed analysis of differences among those observed for a specific number of waves. While the non-absolute value effect size graphs do not display a clear one-sided effect—they display more of a snake-like pattern ebbing and flowing on either side of the zero threshold—the absolute value effect size graphs tell a similar story to that presented in the main text. The largest effects occur when comparing those observed for fewer number of waves and decreases as we compare those observed for extended number of waves.

²³ These means are the averages across all waves the respondents were observed for.

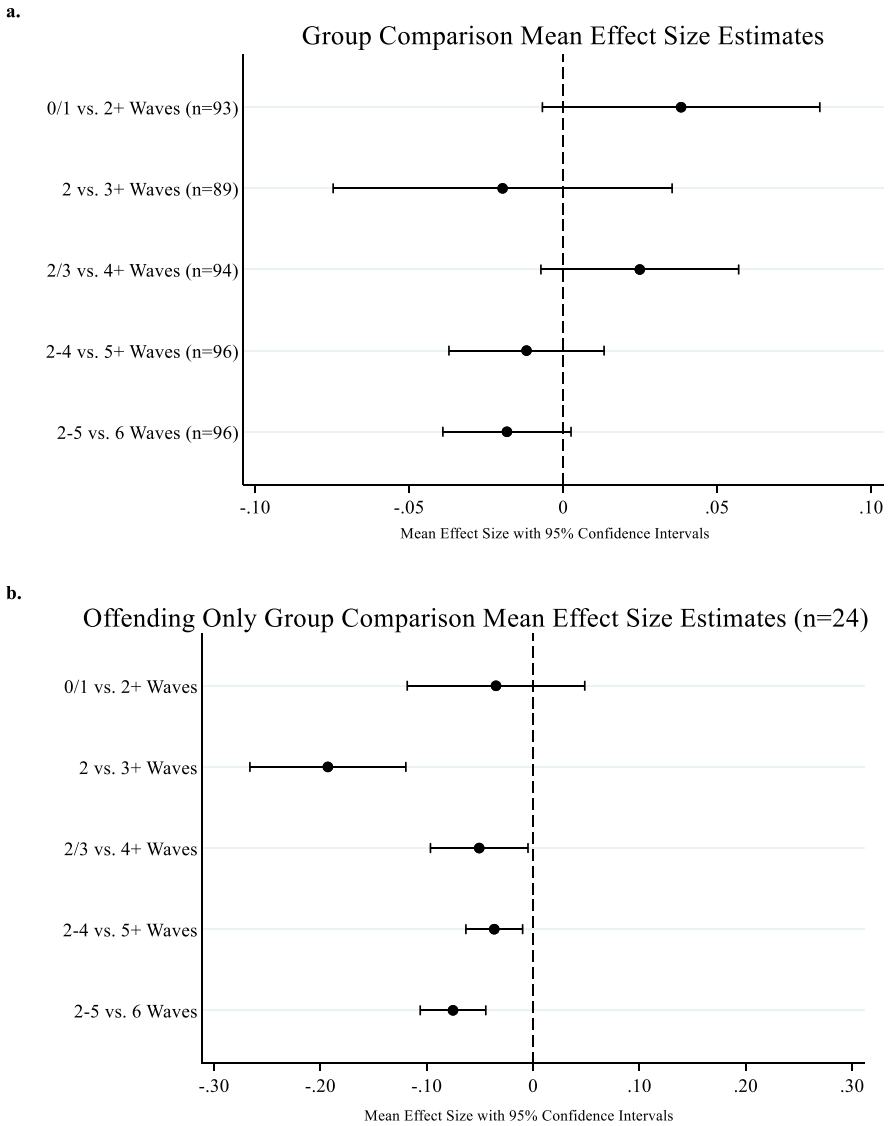


Fig. 1 **a** All variables except offending: group comparison mean effect size estimates with 95% confidence intervals. **b** Offending only: group comparison mean effect size estimates with 95% confidence intervals

require more waves of data, the sample becomes less residentially mobile while having higher levels of perceived social costs. The sample also becomes less criminogenic as we restrict the sample to those who contribute more waves as those who contribute fewer waves are more criminogenic. There are other noticeable changes. Those who contribute fewer waves spend more time in an institution than outside of one. Those who contribute fewer waves also report witnessing more violence,

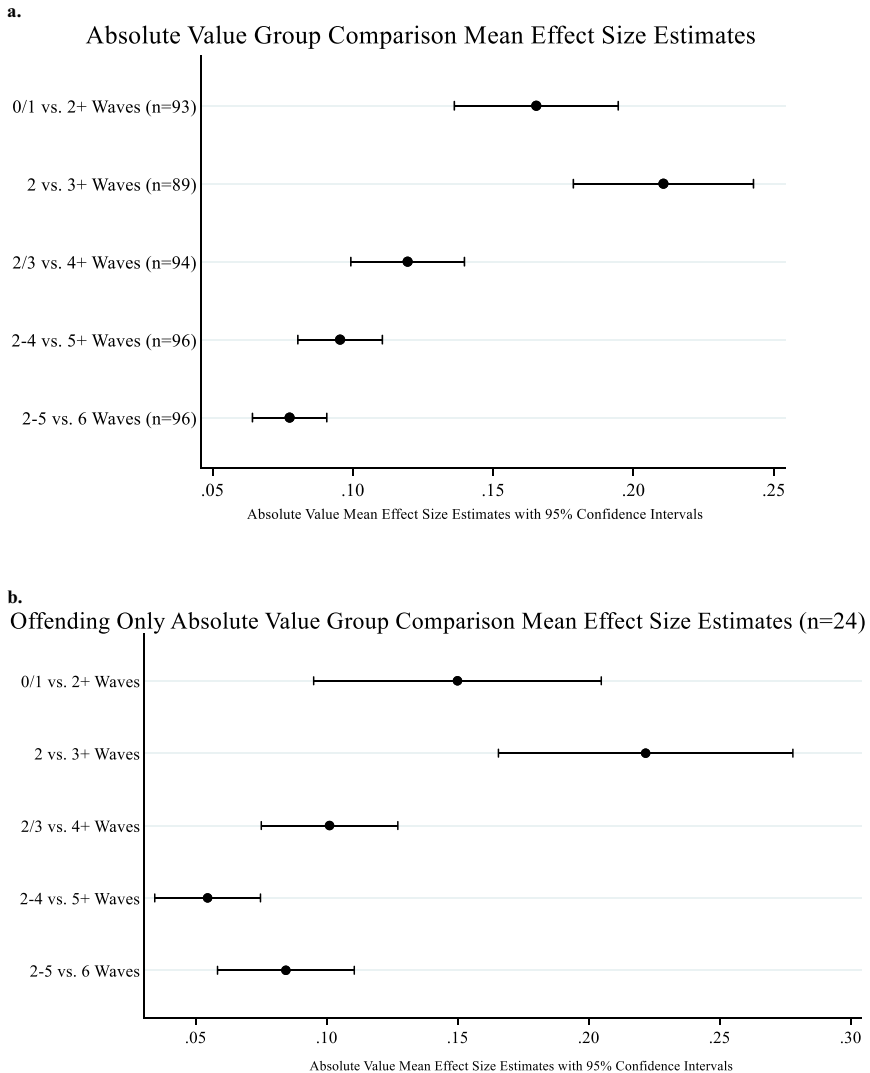


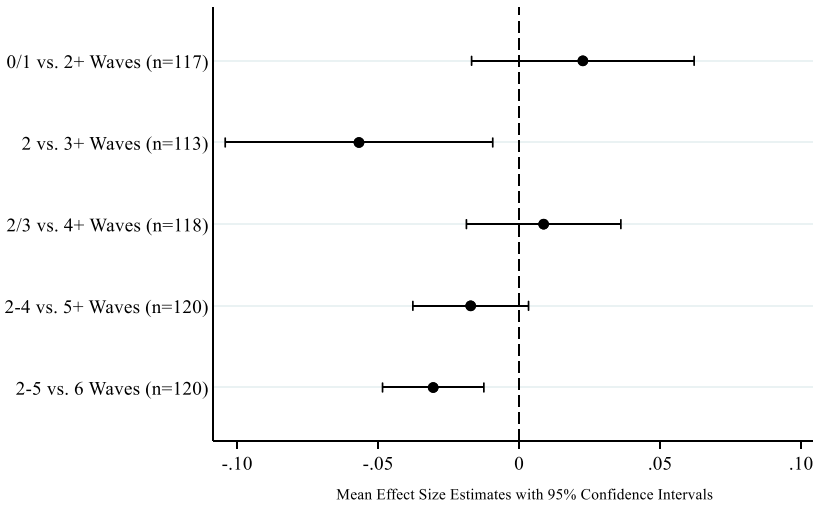
Fig. 2 **a** All variables except offending: absolute value group comparison mean effect size estimates with 95% confidence intervals. **b** Offending only: absolute value group comparison mean effect size estimates with 95% confidence intervals

being victimized more often, and being older on average. Thus, the sample appears to change in important ways as we require more waves of data for our analyses.

We report the results of the fixed effects models across conditioned waves in Table 4. The coefficient for residential mobility has a significant negative relationship with perceived social costs and remains so until we condition on those observed for at least five waves. The coefficient drops in magnitude by 22% as compared to the sample of those observed for at least two waves and is no longer statistically significant.

3a.

All Variables Group Comparison Mean Effect Size Estiamtes



b.

All Variables Absolute Value Group Comparison Mean Effect Size Estiamtes

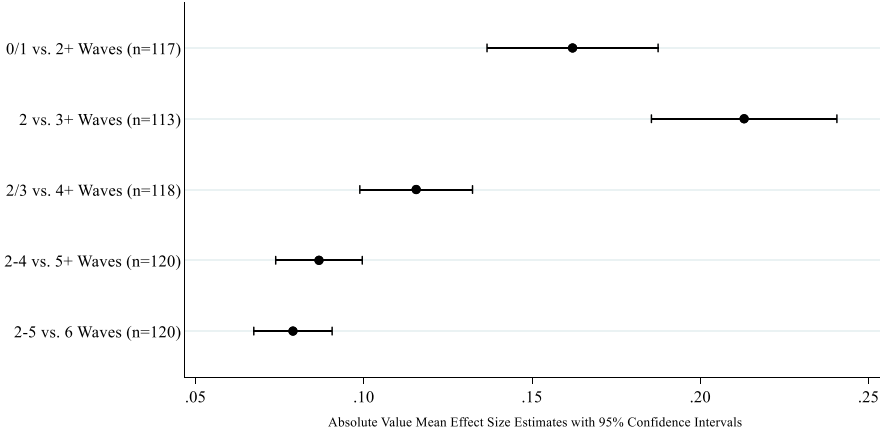


Fig. 3 a All variables: group comparison mean effect size estimates with 95% confidence intervals. b All variables: absolute value group comparison mean effect size estimates with 95% confidence intervals

The coefficient drops by 33%, as compared to the sample of those observed for at least two waves, when conditioning on all six waves and remains insignificant. Furthermore, the coefficient drops by 14% when going from those observed for at least five waves to those observed for all six waves. When looking back at Table 3, the change from significance to insignificance occurred when the mean level of residential mobility took a drop from 0.588 to 0.578 and then to 0.555 residential changes, with corresponding increases in perceived social costs from 3.058 to 3.062 and then

Table 3 Mean levels of variables across conditioned waves

Variables	Conditioned number of waves				
	At least 2 waves	At least 3 waves	At least 4 waves	At least 5 waves	At least 6 waves
	Mean	Mean	Mean	Mean	Mean
Perceived social costs	3.055	3.057	3.058	3.062	3.082
Residential mobility	.589	.589	.588	.578	.555
Offending	1.396	1.395	1.392	1.390	1.341
Age	17.775	17.771	17.764	17.734	17.704
Witnessed violence	.948	.944	.942	.933	.902
Victimized	.190	.190	.188	.184	.184
Antisocial peer influence	1.459	1.459	1.456	1.454	1.457
Ran away	.012	.012	.012	.011	.010
Employed	.532	.531	.533	.534	.545
Gang member	.102	.103	.103	.102	.104
Public housing	.094	.094	.093	.095	.094
Street time	.640	.640	.640	.647	.663
<i>n</i>	1315	1296	1248	1150	885

to 3.082. A similar pattern occurred for offending where it had a significant negative effect until conditioning on those observed for all six waves where it is no longer significant. In comparison to those observed for at least two waves, the offending coefficient dropped by 40% when sampling on those observed for all six waves. The change in significance also matches up to Table 3 where the mean dropped heavily. Thus, the two primary variables of interest—residential mobility and offending—are no longer statistically significant as we required more waves of data to be included in our sample. These results indicate that not only do the characteristics of the sample change when conditioning on different number of waves, but so too does the results.

Discussion

While attrition has been extensively studied in panel datasets across many disciplines, attrition in the singular sense of dropping out and never returning provides only partial insight into a more complex issue specific to panel datasets. An equally important, but understudied, missing data issue in panel datasets is when respondents miss interviews and are re-interviewed at various points throughout the panel dataset. The implication of this dynamic missingness is that the total number of waves respondents contribute to the panel dataset varies across respondents. With the varying number of waves respondents contribute, researchers are faced with an

Table 4 Fixed effects models predicting perceived social costs (clustered SE)

Variables	At least 2 waves <i>b</i> (SE)	At least 3 waves <i>b</i> (SE)	At least 4 waves <i>b</i> (SE)	At least 5 waves <i>b</i> (SE)	All 6 waves <i>b</i> (SE)
Residential mobility	-.036 (.015)*	-.036 (.015)*	-.038 (.015)*	-.028 (.016)	-.024 (.017)
Offending	-.015 (.006)*	-.015 (.006)*	-.015 (.006)*	-.014 (.006)*	-.009 (.007)
Age	.040 (.013)**	.039 (.013)**	.038 (.013)**	.040 (.013)**	.038 (.015)*
Witness violence	.011 (.011)	.010 (.010)	.011 (.011)	.014 (.011)	.014 (.012)
Victimized	.012 (.022)	.012 (.022)	.012 (.022)	.010 (.023)	.004 (.026)
Antisocial peer influence	-.019 (.020)	-.018 (.020)	-.014 (.020)	-.029 (.021)	-.027 (.023)
Ran away	.178 (.126)	.154 (.123)	.128 (.121)	.080 (.128)	.052 (.141)
Employed	.013 (.024)	.014 (.024)	.017 (.024)	.013 (.025)	.006 (.027)
Gang member	.050 (.061)	.050 (.061)	.040 (.061)	.049 (.061)	.067 (.066)
Public housing	.050 (.047)	.051 (.047)	.043 (.047)	.060 (.048)	.070 (.054)
Street time	-.053 (.038)	-.053 (.038)	-.052 (.038)	-.053 (.040)	-.040 (.044)
<i>n</i>	1315	1296	1248	1150	885

SE = standard error

* $p < .05$; ** $p < .01$; *** $p < .001$

important decision before any statistical analyses are conducted: how many interviews will they require respondents to have been observed for to be included in their panel sample. This is not an inconsequential decision given that individuals with characteristics relevant to crime-related inquiries miss more follow-up interviews—especially in panel datasets of individuals with past offending experiences (Western et al., 2016). This is often a subjective decision by the researcher given the extensive degrees of freedom (Gelman & Loken, 2013) they have when weighing the balancing act of wanting more waves to tease out the development effect more precisely against the concern of losing individuals with important crime-related characteristics.

In order to tease this out and demonstrate how this can impact one's sample and subsequent conclusions, we conducted two sets of analyses using the Pathways to Desistance study which is a panel dataset of youth who have previously committed serious offenses. The first set of analyses was centered around *prospectively* assessing what factors were associated with individuals contributing different number of waves in the *future*. Here, we assessed mean differences at the baseline interview across those contributing different number of waves. Not only did offending variables register statistically significant, a number of variables correlated with offending in the criminological literature were also significant. Those contributing fewer waves were older at baseline and entered the peak of the age-crime curve earlier than those observed for more waves. The relationship between age and crime in Western societies has a long scholarly history where crime increases in one's teenage and late teenage years before steadily declining thereafter into adulthood (Hirschi & Gottfredson, 1983; Sampson & Laub, 1993). It is important to include in one's panel

sample those entering the peak of the age-crime curve earlier than others given their increased propensity to offend and difficulty in re-interviewing them. Those interviewed in a facility as compared to the community, and those having more petitions prior to baseline, contributed fewer waves of data. Being interviewed in a facility may be the result of having committed a serious offense as compared to a less serious offense that would allow one to be interviewed outside of a facility. Those with more petitions prior to baseline contributed fewer waves of data, which is indicative of a longer history of offending. Being unable to follow-up on these individuals would result in extended waves of data being comprised of people who have committed fewer serious offenses, or fewer total offenses. Those contributing fewer waves of data also have fewer people in their home, were less likely to have their biological mother at home, were less likely to have living biological parents, and were less likely to attend school prior to their detention. These variables are indicative of these individuals having reduced levels of informal social control as compared to those who contributed more waves of data (Costello & Laub, 2020; Hirschi, 1969; Sampson & Laub, 1993; Toby, 1957). These are also variables that tap into sources of social support that extend beyond psychological variables that Western (2018) identified as a hindrance to re-interviewing hard-to-reach respondents.

While these variables were statistically significant and important in their own right as they are often correlated with offending, we also examined effect sizes to better understand the magnitude of the effects of these sampling processes. Aggregating up to the group-level the individual variable effect sizes, we found that the effect sizes of these differences were quite large and indicative of meaningful differences across samples. Those who contributed fewer waves were different on these baseline characteristics than those contributing more waves. This was especially noticeable for the offending variables, and the absolute value of the mean effect sizes for both offending and non-offending variables. Importantly, the largest differences in effect sizes were seen when comparing those who contributed the fewest waves of data for panel data analyses (2 waves) against those who contributed more waves of data (3+ waves). This finding extends beyond individual variables to show that, at the sample level, those contributing fewer waves contain fundamentally important factors that may bias one's sample should they condition on an extended number of waves.

The second set of analyses was an empirical test examining the long-studied relationship between residential mobility and informal social costs. Similar to other processes when studying crime—particularly for life-course and developmental studies—the key variables of interest are likely affected by those who differentially miss interviews. As compared to the first set of analyses, this approach mixes the prospective sampling differences with time-varying factors on the relationship under study. Employing a fixed effects model that removes all time-stable influences, as we progressively required more waves of respondents to be entered in our sample, the sample became less residentially mobile with higher levels of informal social costs, making the initial statistically significant negative relationship insignificant. A similar trend was identified for the relationship between offending and informal social costs. Thus, the characteristics of the sample and results were sensitive to the number of waves we conditioned our sample upon. These findings couple those comparing the means of the variables across waves.

These findings speak directly to all scholars using panel datasets, and the many researcher degrees of freedom available to them (Gelman & Loken, 2013; Simmons et al., 2011). Although we demonstrated how the sample and conclusions change with the relationship between residential mobility and informal costs, most if not all scholars employing panel and developmental methods are subject to the balancing act of wanting more waves against the concern of losing those who contain unique characteristics. Two commonly studied processes within criminology, among many others, illustrate this balancing act and how scholars can use our findings to inform their research.

The first example concerns the long-debated relationship between desistance from crime arising from bonds to marriage and employment (Sampson & Laub, 1993), and whether one's identity to cease offending precedes commitment to marriage and employment and subsequent reductions in offending (Gottfredson & Hirschi, 2020; Paternoster & Bushway, 2009; Paternoster et al., 2015). Not only do scholars likely need many waves of data, but they also need many waves of data on those who have offended specifically such as the dataset we used in this study. To tease out this process more precisely, many consecutive waves of data would likely be desired in order to capture differences as to when respondents change in their identities which *then* starts the desistance process of becoming attached to marriage and work which *then* leads to reduced offending. Our results inform at least two cautionary steps researchers should consider when studying such processes. First, there may be meaningful differences at the baseline interview that are *prospectively* related to reporting different number of waves. This would uncover variables, both those under study and other variables that are not under study but related to offending, that differentially influence those who are observed for differing number of waves. Such differences may include not just past offending, but potential labeling one received after their offending. One may identify with that criminal label, and those identifying with that label may be differentially represented across conditioned samples. Second, those who have the strongest deviant identities, are the least tied to informal institutions, and offend the most may be the hardest to follow-up with and thus be unable to contribute to extended waves of data. Those able to contribute would be more pro-social and possibility distort conclusions as they are fundamentally different from those who did not contribute the required waves.

The second example pertains to the deleterious effects of being incarcerated as a youth. While it is typically agreed upon that being incarcerated as a youth increases one's future offending (Nagin et al., 2009), scholars have expanded their inquiries to assess how long into one's adult life these effects last (Kurlychek et al., 2022). As scholars build on this research to explain the specific mechanisms incarceration operates through to increase one's offending, and its impact on turning points toward desistance, rich survey data is needed pertaining to the personal, psychological, and sociological characteristics of the respondents. There may be important baseline differences that are *prospectively* associated with how many waves one contributes. These may include past institutionalization, and also differential deleterious effects felt by the observed youth due to the institutionalization where those most negatively affected may be hardest to follow-up with. By increasing the number of waves respondents were required to have been observed for, scholars may lose those who

are most criminally active who also have personal, psychological, and sociological disadvantages. This would affect results appraising the mechanism institutionalization operates through and depict a process that is not applicable to the most criminogenic respondents.

Our results also speak to panel data collection efforts focused on individuals with experiences of past offending. While it is commonly accepted that individuals who have offended, and those who have committed more serious offenses, are difficult to follow up on, Western (2018) and colleagues (2016) have extensively detailed these issues and the importance of dedicating effort to follow up on these individuals (see also Clark et al., 2020; Fahmy et al., 2019). Indeed, our study examining youth who have committed serious offenses demonstrated similar processes wherein those who offended the most contributed fewer waves—even though the study investigators invested much effort to follow up on these youth. If researchers are collecting their own data, they should attend to current and best practices for ensuring maximum follow-up retention rates at all waves (Clark et al., 2020; Fahmy et al., 2019; Western, 2018). If researchers are using existing data, they should heed our findings and follow our recommendations and associated flowchart we discuss in the upcoming paragraphs.

We leverage two broad suggestions from scholars in interdisciplinary fields to propose clear steps researchers studying crime can take to help diagnose and combat these issues. Jeličić et al., (2009, p. 1199) propose that researchers clearly report the number of waves they condition their panel sample upon and also how the sample characteristics of the conditioned sample differs from the baseline sample. Ware et al. (2012), as an expert panel of the National Research Council (NRC), suggested sensitivity analyses of one's chosen sample against other potential samples. While these recommendations are useful, and we endorse them, we also propose concrete steps scholars should take when appraising panel datasets. Our recommendations are not cures or fixes to missed waves of data. They are, however, diagnostics and ways to help scholars make more informed decisions about their sample.

We propose that scholars examine their relationship under inquiry by comparing a model with the fewest number of waves available to answer their theoretical question against a statistical model using all waves of data that are available. In our case, this entails those observed for at least two waves against those observed for all six waves.²⁴ In doing so, researchers should look for two things. The first is if the coefficients for the primary variables of interest change by 20% or more. The second is to check if the coefficients for these variables change in statistical significance. A change in statistical significance clearly indicates an issue across waves due to lost participants. A change in the coefficient of 20% or more, which is the same value used for the standardized bias value in propensity score matching (Angrist &

²⁴ This is important to note because different models have different number of waves that constitute the fewest number of waves the model can analyze. For fixed effects models, one would examine the change in coefficients between those observed for at least two waves against those observed for all of the waves. For other statistical models examining growth—such as growth curve models or group-based trajectory models—the fewest waves to evaluate non-linear change is three waves.

Pischke, 2009; Rosenbaum & Rubin, 1985), is a useful proxy for identifying issues. In our empirical demonstration, the primary variables of interest changed in statistical significance when the change in the coefficients were all greater than 20% with the smallest of these changes being a 22% change.

If the percent change in the primary coefficients of interest are less than 20% *and* the coefficients do not change in statistical significance, it seems safe to say that one's model is resilient to the effects of missed interviews. However, if the percent change in the primary coefficients are greater than 20% *and/or* the coefficients change in statistical significance, we encourage researchers to not only report these changes, but also disclose the direction of the change and further assess what baseline characteristics may be impacting the results.²⁵ In this instance, using a statistical model that requires fewer waves of data is recommended as fewer individuals with important characteristics are missing. In Fig. 4, we provide a flowchart for applying our recommendations.

Our recommendations also speak to those employing common missing data techniques, such as full information maximum likelihood (FIML; Allison, 2002; see also Allison, 2015, for a good description on the utility of FIML). After one conducts our above-recommended steps, they should also do the same steps with models employing FIML. If the missing data is either MCAR or MAR, we would expect the percent change in the coefficients to be smaller when employing FIML. This would indicate that FIML is able to overcome at least some of the bias arising from the missing data. This would be especially useful if, after employing FIML, the percent change in coefficients is less than 20% *and* the coefficients do not change in statistical significance. However, if the percent change in coefficients persists when using FIML as opposed to not using FIML, that would be evidence of bias that is unable to be overcome where the missing data are missing not at random (MNAR; Allison, 2002).

We conclude by echoing Allison's (2002) comment that the best solution to missing data is not to have any. However, missing data will occur. We demonstrated one way that has thus far been overlooked: missed interviews and the consequences of conditioning one's sample on those observed for different number of waves. We hope we have provided researchers with the tools to more critically appraise their sample and results without uncritical reliance on statistical methods with herculean assumptions. Panel data studies should continue unabated, but researchers should spend more time understanding their sample and the factors related to being observed for different number of waves.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40865-022-00207-w>.

²⁵ Although it would be desirable to provide a statistical test to make these decisions easier, a formal statistical test is complicated by that fact that the conditioned samples are not independent of one another (see Paternoster et al., 1998). That is, those who are observed for at least two waves may also be included in a sample of those observed for all six waves.

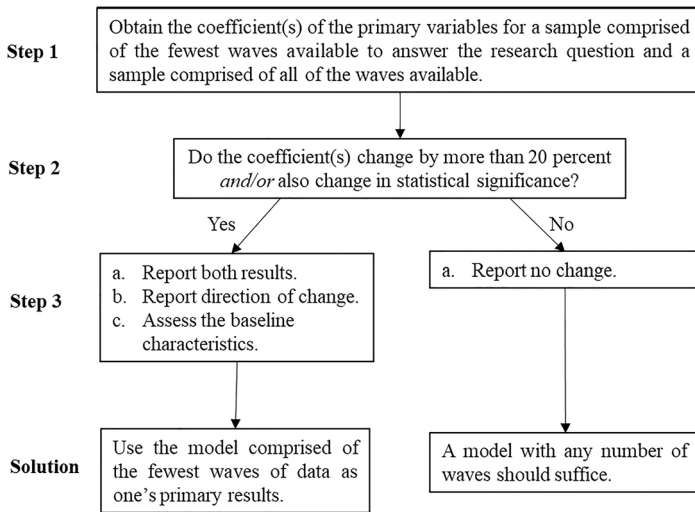


Fig. 4 Flowchart for applying our recommendations

Acknowledgements We would like to thank Stefan Ivanov, David McDowall, Manuel Eisner, and the anonymous reviewers for providing us with valuable feedback on earlier versions of this paper. This paper was presented at the Annual Meeting of the American Society of Criminology (ASC) in Chicago in 2021. We thank the ASC attendees who provided us with feedback.

Declarations

Conflict of Interest The authors declare no competing interests.

References

- Allison, P. D. (2002). *Missing data*. Sage.
- Allison, P. D. (2009). *Fixed effects regression models*. Sage.
- Allison, P. D. (2015). Maximum likelihood is better than multiple imputation: Part II. Retrieved from <https://statisticalhorizons.com/ml-is-better-than-mi/>
- Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- Badawi, M. A., Eaton, W. W., Myllyluoma, J., Weimer, L. G., & Gallo, J. (1999). Psychopathology and attrition in the Baltimore ECA 15-year follow-up 1981–1996. *Social Psychiatry Psychiatric Epidemiology*, *34*(2), 91–98.
- Beaver, K. M. (2013). Intelligence and selective attrition in a nationally representative and longitudinal sample of Americans. *Personality and Individual Differences*, *55*, 157–161.
- Berzofsky, M. E., & Carrillo-Garcia, I. (2018). Determining the optimal number of interview waves in a panel survey with application to the National Crime Victimization Survey. *Science Journal of Applied Mathematics and Statistics*, *6*(5), 135–147.
- Blumstein, A. (1983). Selective incapacitation as a means of crime control. *American Behavioral Scientist*, *27*(1), 87–108.
- Blumstein, A., Cohen, J., & Nagin, D. (1978). *Deterrence and incapacitation: Estimating the effects of sanctions on the crime rate*. National Academy Press.

- Blumstein, A., & Cohen, J. (1979). Estimation of individual crime rates from arrest records. *The Journal of Criminal Law & Criminology*, 70(4), 561–585.
- Blumstein, A., Cohen, J., & Farrington, D. P. (1988). Criminal career research: Its value for criminology. *Criminology*, 26(1), 1–35.
- Blumstein, A., Cohen, J., & Hsieh, P. (1982). *The duration of adult criminal careers. Final report to the National Institute of Justice*. Urban Systems Institute.
- Blumstein, A., Cohen, J., Roth, J., & Visser, C. (1986). *Criminal careers and "career criminals."* National Academy Press.
- Boys, A., Marsden, J., Stillwell, G., Hatchings, K., Griffiths, P., & Farrell, M. (2003). Minimizing respondent attrition in longitudinal research: Practical implications from a cohort study of adolescent drinking. *Journal of Adolescence*, 26(3), 363–373.
- Caruana, E. J., Roman, M., Hernández-Sánchez, J., & Solli, P. (2015). Longitudinal studies. *Journal of Thoracic Disease*, 7(11), E537–E540.
- Clark, K. J., Mitchell, M. M., Fahmy, C., Pyrooz, D. C., & Decker, S. H. (2020). What if they are all high-risk for attrition? Correlates of retention in a longitudinal study of reentry from prison. *International Journal of Offender Therapy and Comparative Criminology*. <https://doi.org/10.1177/0306624X20967934>
- Cohen, J. (1984). Selective incapacitation: An assessment. *University of Illinois Law Review*, 2, 253–290.
- Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). Lawrence Erlbaum Associates.
- Cordray, S., & Polk, K. (1983). The implications of respondent loss in panels studies of deviant behavior. *Journal of Research in Crime and Delinquency*, 20(2), 214–242.
- Costello, B. J., & Laub, J. H. (2020). Social control theory: The legacy of Travis Hirschi's causes of delinquency. *Annual Review of Criminology*, 3, 21–41.
- Cotter, R. B., Burke, J. D., Loeber, R., & Navratil, J. L. (2002). Innovative retention methods in longitudinal research: A case study of the developmental trends study. *Journal of Child and Family Studies*, 11(2), 485–498.
- Cotter, R. B., Burke, J. D., Stouthamer-Loeber, M., & Loeber, R. (2005). Contacting participants for follow-up: How much effort is required to retain participants in longitudinal studies? *Evaluation and Program Planning*, 28(1), 15–21.
- Cullen, F. T., Pratt, T. C., & Graham, A. (2019). Why longitudinal research is hurting criminology. *The Criminologist*, 44(2), 2–7.
- Cunradi, C. B., Moore, R., Killoran, M., & Ames, G. (2005). Survey nonresponse bias among young adults: The role of alcohol, tobacco, and drugs. *Substance Use and Misuse*, 40(2), 171–185.
- de Graaf, R., Bijl, R. V., Smit, F., Ravelli, A., & Vollebergh, W. A. M. (2000). Psychiatric and sociodemographic predictors of attrition in a longitudinal study The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *American Journal of Epidemiology*, 11(1), 1039–1047.
- Delfabbro, P., Winefield, H., Winefield, A., Malvaso, C., & Plueckhahn, T. (2017). Factors associated with attrition in a 10-year longitudinal study of young people: Implications for studies of employment in school leavers. *Australian Psychologist*, 52(1), 41–51.
- Eaton, W. W., Anthony, J. C., Tepper, S., & Dryman, A. (1992). Psychopathology and attrition in the epidemiologic catchment area surveys. *American Journal of Epidemiology*, 135(9), 1051–1059.
- Eidson, J. L., Roman, C. G., & Cahill, M. (2017). Successes and challenges in recruiting and retaining gang members in longitudinal research: Lessons learned from a multisite social network study. *Youth Violence and Juvenile Justice*, 15(4), 396–418.
- Eisner, N. L., Murray, A. L., Eisner, M., & Ribeaud, D. (2019). A practical guide to the analysis of non-response and attrition in longitudinal research using a real data example. *International Journal of Behavioral Development*, 43(1), 24–34.
- Elder, G. H., & Giele, J. Z. (2009). Life course studies: An evolving field. In G. H. Elder & J. Z. Giele (Eds.), *The craft of life course research* (pp. 1–24). The Guilford Press.
- Erickson, M. L. (1973). Delinquency in a birth cohort: A new direction in criminological research. *Journal of Criminal Law and Criminology*, 64(3), 362–367.
- Fahmy, C., Clark, K. J., Mitchell, M. M., Decker, S. H., & Pyrooz, D. C. (2019). Method to the madness: Tracking and interviewing respondents in a longitudinal study of prisoner reentry. *Sociological Methods & Research*, 51(3), 1282–1324.
- Farrington, D. P., Gallagher, B., Morley, L., St. Ledger, R. J., & West, D. J. (1990). Minimizing attrition in longitudinal research: Methods of tracing and securing cooperation in a 24-year follow-up study.

- In D. Magnusson & L. Bergman (Eds.), *Data quality in longitudinal research*. Cambridge University Press.
- Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An analysis of sample attrition in panel data: The Michigan panel study of income dynamics. *The Journal of Human Resources*, *33*(2), 251–299.
- Fitzgerald, J. M. (2011). Attrition in models of intergenerational links using the PSID with extensions to health and sibling models. *B E Journal of Economic Analysis & Policy*, *11*(3), 1–58.
- Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf
- Glueck, S., & Glueck, E. T. (1930). *500 criminal careers*. Knopf.
- Glueck, S., & Glueck, E. T. (1950). *Unraveling juvenile delinquency*. Harvard University Press.
- Glueck, S., & Glueck, E. T. (1968). *Delinquents and nondelinquents in perspective*. Harvard University Press.
- Gottfredson, M., & Hirschi, T. (1986). The true value of lambda would appear to be zero: An essay on career criminals, criminal careers, selective incapacitation, cohort studies, and related topics. *Criminology*, *24*(2), 213–234.
- Gottfredson, M., & Hirschi, T. (1987). The methodological adequacy of longitudinal research on crime. *Criminology*, *25*(3), 581–614.
- Gottfredson, M., & Hirschi, T. (1988). Science, public policy, and the career paradigm. *Criminology*, *26*(1), 37–55.
- Gottfredson, M. R. & Hirschi, T. (2020). *Modern control theory and the limits of criminal justice*. Oxford University Press.
- Gustavson, K., von Soest, T., Karevold, E., & Røysamb, E. (2012). Attrition and generalizability in longitudinal studies: Findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC Public Health*, *12*, 918.
- Hathaway, S., & Monachesi, E. (1953). *Analyzing and predicting juvenile delinquency with the MMPI*. University of Minnesota Press.
- Hauser, R. M. (2005). Survey response in the long run: The Wisconsin longitudinal study. *Field Methods*, *17*(1), 3–29.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251–1271.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.
- Hirschi, T. (1969). *Causes of delinquency*. University of California Press.
- Hirschi, T., & Gottfredson, M. (1983). Age and the explanation of crime. *American Journal of Sociology*, *89*(3), 552–584.
- Huesmann, L. R., Eron, L. D., Lefkowitz, M. M., & Walder, L. O. (1984). The stability of aggression over time and generations. *Developmental Psychology*, *20*(6), 1120–1134.
- Horney, J. D., Osgood, W., & Marshall, I. H. (1995). Criminal careers in the shortterm: Intra-individual variability in crime and its relation to local life circumstances. *American Sociological Review*, *60*(5), 655–673.
- Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*(4), 1195–1199.
- Joshi, H. (2016). Why do we need longitudinal survey data? Retrieved from <https://wol.iza.org/articles/why-do-we-need-longitudinal-survey-data/long>
- Keyes, K. M., Jager, J., Platt, J., Rutherford, C., Patrick, M. E., Kloska, D. D., & Schulenberg, J. (2020). When does attrition lead to biased estimates of alcohol consumption? Bias analysis for loss to follow-up in 30 longitudinal cohorts. *International Journal of Methods in Psychiatric Research*, *29*, 1–9.
- Kijowski, M. C., & Wilson, T. (2021) Examining the effect of residential mobility on crime through the mechanism of perceived informal social costs. *Crime and Delinquency*. <https://doi.org/10.1177/00111287211010485>.
- Kirk, D. S. (2020). *Home free: Prisoner reentry and residential change after hurricane Katrina*. Oxford University Press.
- Kurlychek, M. C., Kijowski, M. C., & Gagnon, A. M. (2022). The long-term consequences of imprisoning our youth: The lasting impact of time spent in adult jails and prisons. *Social Problems*. <https://doi.org/10.1093/socpro/spab078>.

- Launes, J., Hokkanen, L., Laasonen, M., Tuulio-Henriksson, A., Virta, M., Lipsanen, J., Tienari, P. J., & Michelsson, K. (2014). Attrition in a 30-year follow-up of a perinatal birth risk cohort: Factors change with age. *PeerJ*. <https://doi.org/10.7717/peerj.480>
- Lugtig, P. (2014). Panel attrition: Separating stayers, fast attriters, gradual attriters, and lurkers. *Sociological Methods & Research*, 43(4), 699–723.
- Maruna, S. (2001). *Making good: How ex-offenders reform and reclaim their lives*. American Psychological Association Books.
- McCord, W. M., & McCord, J. (1959). *Origins of crime: A new evaluation of the Cambridge-Somerville youth study*. Columbia University Press.
- Mitchell, M. M., Fahmy, C., Clark, J. K., & Pyrooz, D. C. (2021). Non-random study attrition: Assessing correction techniques and the magnitude of bias in a longitudinal study of reentry from prison. *Journal of Quantitative Criminology*. <https://doi.org/10.1007/s10940-021-09516-7>.
- Mulvey, E. P. (2012). Research on pathways to desistance [Maricopa County, AZ and Philadelphia County, PA]: Subject measures, 2000–2010. ICPSR29961-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research
- National Research Council (2014). *The growth of incarceration in the United States: Exploring causes and consequences*. In J. Travis, B. Western, & F.S Redburn (Eds.), The National Academies Press.
- Nagin, D. S., Cullen, F. T., & Jonson, C. L. (2009). Imprisonment and reoffending. *Crime and Justice*, 38(1), 115–200.
- Nagin, D. S., & Paternoster, R. (1994). Personal capital and social control: The deterrence implications of a theory of individual differences in criminal offending. *Criminology*, 32(4), 581–606.
- Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2011). A meta-analysis of predictors of offender treatment attrition and its relationship to recidivism. *Journal of Consulting and Clinical Psychology*, 79(1), 6–21.
- Pager, D. (2007). *Marked: Race, crime and finding work in an era of mass incarceration*. University of Chicago Press.
- Paternoster, R., & Bushway, S. (2009). Desistance and the “feared self”: Toward an identity theory of criminal desistance. *The Journal of Criminal Law & Criminology*, 99(4), 1103–1156.
- Paternoster, R., Bachman, R., Bushway, S., Kerrison, E., & O’Connell, D. (2015). Human agency and explanations of criminal desistance: Arguments for a rational choice theory. *Journal of Developmental and Life-Course Criminology*, 1, 209–235.
- Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36(4), 859–866.
- Pathways to Desistance (n.d.) Baseline interview content. Retrieved from <https://www.pathwaysstudy.pitt.edu/baselineinterview.html>.
- Pettit, B. (2012). *Invisible men: Mass incarceration and the myth of Black progress*. Russell Sage Foundation.
- Powers, E., & Witmer, H. L. (1951). *An experiment in the prevention of delinquency: The Cambridge-Somerville youth study*. Columbia University Press.
- Reinhart, A. (2015). *Statistics done wrong: The woefully complete guide*. No Starch Press.
- Remster, B. (2019). A life course analysis of homeless shelter use among the formerly incarcerated. *Justice Quarterly*, 36(3), 437–465.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33–38.
- Saiepour, N., Ware, R., Najman, J., Baker, P., Clavarino, A., & Williams, G. (2016). Do participants with different patterns of loss to follow-up have different characteristics? A multi-wave longitudinal study. *Journal of Epidemiology*, 26(1), 45–49.
- Sampson, R. J. (1991). Linking the micro- and macrolevel dimensions of community social organization. *Social Forces*, 70(1), 43–64.
- Sampson, R. J., & Laub, J. H. (1993). *Crime in the making: Pathways and turning points through life*. Harvard University Press.
- Sampson, R. J., & Laub, J. H. (2016). Turning points and the future of life-course criminology: Reflections on the 1986 criminal careers report. *Journal of Research in Crime and Delinquency*, 53(3), 321–335.
- Satherley, N., Milojev, P., Greaves, L. M., Huang, Y., Osborne, D., Bulbulia, J., & Sibley, C. G. (2015). Demographic and psychological predictors of panel attrition: Evidence from the New Zealand attitudes and values study. *PLoS One*, 10(3), 1–22.

- Schoeni, R. F., & Wiemers, E. E. (2015). The implications of selective attrition for estimates of intergenerational elasticity of family income. *Journal of Economic Inequality*, *13*, 351–372.
- Schubert, C. A., Mulvey, E. P., Steinberg, L., Cauffman, E., Losoya, S. H., Hecker, T., Chassin, L., & Knight, G. P. (2004). Operational lessons from the Pathways to Desistance project. *Youth Violence and Juvenile Justice*, *2*(3), 237–255.
- Shaw, C. R., & McKay, H. D. (1942). *Juvenile delinquency and urban areas*. University of Chicago Press.
- Siddiqui, O., Flay, B. R., & Hu, F. B. (1996). Factors affecting attrition in a longitudinal smoking prevention study. *Preventive Medicine*, *25*(5), 554–560.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712.
- Sugie, N. F., Zatz, N. D., & Augustine, D. (2020). Employer aversion to criminal records: An experimental study of mechanisms. *Criminology*, *58*(1), 5–34.
- Sweeten, G. (2012). Scaling criminal offending. *Journal of Quantitative Criminology*, *28*, 533–557.
- Thimasarn-Anwar, T., Sanders, J., Munford, R., Jones, G., & Liebenberg, L. (2014). Rethinking late and lost to follow-up participants: The New Zealand youth transitions study. *Journal of Youth Studies*, *17*(5), 626–641.
- Thornberry, T. P., Bjerregaard, B., & Miles, W. (1993). The consequences of respondent attrition in panel studies: A simulation based on the Rochester Youth Development Study. *Journal of Quantitative Criminology*, *9*(2), 127–158.
- Thornberry, T. P., & Krohn, M. D. (2003). The development of panel studies of delinquency. In T. Thornberry & M. Krohn (Eds.), *Taking stock of delinquency: An overview of findings from contemporary longitudinal studies* (pp. 1–9). Kluwer Academic/Plenum Publishers.
- Thygesen, L. C., Johansen, C., Keiding, N., Giovannucci, E., & Grønbaek, M. (2008). Effects of sample attrition in a longitudinal study of the association between alcohol intake and all-cause mortality. *Addiction*, *103*(7), 1149–1159.
- Toby, J. (1957). Social disorganization and stake in conformity: Complementary factors in the predatory behavior of hoodlums. *The Journal of Criminal Law, Criminology, and Police Science*, *48*(1), 12–17.
- Vogel, M., Porter, L. C., & McCuddy, T. (2017). Hypermobility, destination effects, and delinquency: Specifying the link between residential mobility and offending. *Social Forces*, *95*(3), 1261–1284.
- Ware, R. S., Williams, G. M., & Aird, R. L. (2006). Participants who left a multiple-wave cohort study had similar baseline characteristics to participants who returned. *Annals of Epidemiology*, *16*(11), 820–823.
- Ware, J. H., Harrington, D., Hunter, D. J., & D'Agostino, R. B. (2012). Missing data. *The New England Journal of Medicine*, *367*, 1353–1354.
- Warner, B. D., & Rountree, P. W. (1997). Local social ties in a community and crime model: Questioning the systemic nature of informal social control. *Social Problems*, *44*(4), 520–536.
- Watkins, A. M., & Melde, C. (2007). The effect of self-control on unit and item nonresponse in an adolescent sample. *Journal of Research in Crime and Delinquency*, *44*(3), 267–294.
- Weinberger, D. A., Tublin, S. K., Ford, M. E., & Feldman, S. S. (1990). Preadolescents' social-emotional adjustment and selective attrition in family research. *Child Development*, *61*(5), 1374–1386.
- West, D. J., & Farrington, D. P. (1973). *Who becomes delinquent: Second report on the Cambridge Study in delinquent development*. Heinemann.
- Western, B. (2018). *Homeward: Life in the year after prison*. Russell Sage Foundation.
- Western, B., Braga, A., & Kohl, R. (2017). A longitudinal survey of newly-released prisoners: Methods and design of the Boston Reentry Study. *Federal Probation*, *81*(1), 32–40.
- Western, B., Braga, A., Hureau, D., & Sirois, C. (2016). Study retention as bias reduction in a hard-to-reach population. *Proceedings of the National Academy of Sciences*, *113*(20), 5477–5485.
- Widdowson, A. O., & Siennick, S. E. (2021). The effects of residential mobility on criminal persistence and desistance during the transition to adulthood. *Journal of Research in Crime and Delinquency*, *58*(2), 151–191.
- Winship, C., & Zhuo, X. (2018). Interpreting t-statistics under publication bias: Rough rules of thumb. *Journal of Quantitative Criminology*, *36*, 329–346.

- Wolfgang, M. E., Figlio, R. M., & Sellin, T. (1972). *Delinquency in a birth cohort*. University of Chicago Press.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.
- Young, A. F., Powers, J. R., & Bell, S. L. (2006). Attrition in longitudinal studies: Who do you lose? *Australian and New Zealand Journal of Public Health*, 30(4), 353–361.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.