electronic ***International Journal of Time Use Research***
2016, Vol. 13, No. 1, 1- 33 dx.doi.org/10.13085/eIJTUR.13.1.1-33

Measuring short and rare activities – Time diaries in criminology

Sjoerd Van Halem, Evelien M. Hoeben, Wim Bernasco and Tom F. M. Ter Bogt

Sjoerd Van Halem (for correspondence)
Staff Social and Behavioural Sciences
Utrecht University
Domplein 29, 3512 JE Utrecht, Netherlands
E-mail: s.vanhalem@uu.nl

Evelien M. Hoeben
Netherlands Institute for the Study of Crime and Law Enforcement
University of Amsterdam
1012 WX Amsterdam, Netherlands
E-mail: e.m.hoeben@uva.nl

Wim Bernasco
Netherlands Institute for the Study of Crime and Law Enforcement
University of Amsterdam
1012 WX Amsterdam, Netherlands
E-mail: wbernasco@nscr.nl

Tom F. M. Ter Bogt
Staff Social and Behavioural Sciences
Utrecht University
Domplein 29, 3512 JE Utrecht, Netherlands
E-mail: t.f.m.terbogt@uu.nl

Abstract

Motivated by recent time use studies in criminology, this study examined whether time diaries are suitable for measuring short and rare activities such as offending. The study compared time diary data collected among 843 adolescents from the conurbation of The Hague (the Netherlands) with stylized questionnaire data from the same respondents, and with stylized questionnaire data from another sample that is representative for Dutch adolescents (N = 1849). Based on the reported offenses in the diaries (N = 101), findings indicate that time diaries may underestimate population offense rates and may not capture offenses committed by low-frequent offenders. On the other hand, time diaries seem able to measure changes in individuals' involvement in offending over time and to capture most of the situational conditions under which offenses occur. The study concludes with suggestions for dealing with the problems associated with measuring short and rare activities.

JEL-Codes: C00, C80, C83, K42

Keywords: Short and rare activities, crime, time diaries, validity, time use methodology

We wish to thank Per-Olof Wikström for developing the questionnaire and the Space-Time Budget interview in the Peterborough Adolescent and Young Adult Development Study (PADS+) and for sharing this material with us. Beth Hardie, Caroline Moul and Neema Trivedi helped train our interview staff, which we are very grateful for. We would like to thank Kirsten Grandia and Lieneke Spel for taking part in managing the fieldwork. The Research and Documentation Centre (WODC) of the Netherlands' Ministry of Security and Justice made available the data of the Youth Delinquency Survey (MZJ). The WODC cannot be held responsible for the completeness, correctness, and use of this data.

1 Introduction

The time diary is a useful tool for collecting detailed information on activity patterns and individual lifestyles, and often the preferred method to study a variety of research questions across several disciplines (Pentland, Harvey, Lawton and McColl, 1999). However, it has been suggested that time diaries suffer from validity problems when measuring short and rare activities (Frazis and Stewart, 2012; Gershuny 2011; Gershuny, 2012; Juster, 1985; Phipps and Vernon, 2009). These problems concern the study of activities that are rare but not short (e.g., bowling or visiting the theater), as well as activities that are short but not rare (e.g., going to the restroom). Knowledge of the potential limitations of time diaries and how to deal with them might be particularly helpful for time use researchers who predominantly study phenomena that are both short and rare. Criminology is a case in point, as its main focus is on offending: An activity that is both rare and of short duration.

Time diaries have recently been introduced in criminology (Wikström and Butterworth 2006; Wikström, Ceccato, Hardie and Treiber, 2010; Wikström, Oberwittler, Treiber and Hardie, 2012a). Wikström and colleagues developed a new time use method that also captured the geographical location of respondents, and that is specifically concerned with measuring offending, victimization, and other criminologically relevant activities: the Space-Time Budget method (Wikström, et al., 2010; Wikström, et al., 2012a; Wikström, Treiber and Hardie, 2012b). The Space-Time Budget method (STB) has great potential for criminological research. For example, because it enables an accurate operationalization of 'risky' lifestyles (e.g., Osgood et al., 1996), because it enables the investigation of situations that potential offenders are exposed to, and because it directly measures the conditions under which offending and victimization occur (Averdijk and Bernasco, 2014; Bernasco, Ruiter, Bruinsma, Pauwels and Weerman, 2013). Since the method captures not only the activity, but multiple domains of respondents' time use (e.g., where it occurred, who else was present, secondary activities), the method allows for analyzing activities in a more conceptually relevant manner.

For most people, offending is a rare occasion (Van der Laan and Blom 2006). Per illustration, Wikström et al. (2012a) found that 49 percent of their sample reported less than ten acts of offending across five years. Offending is also generally assumed to be of short duration[1]. For example, Buckle and Farrington (1984) found in an observational study that shoplifters spend on average 11 minutes in the store. As offending is rare and assumed to be of short duration, it is unclear whether time diaries can be used to capture the distribution of offending across individuals or the development of involvement in offending over time. The present study examines these and related questions by comparing time diary data that was collected in the Study of Peers, Activities and Neighborhoods (SPAN), with two datasets: A stylized questionnaire

[1] This might vary for different types of offending.

administered to the same respondents (also collected in the SPAN study), and external data of the Youth Delinquency Survey (MZJ). The latter contains information on offending from a sample that is representative of adolescents in the Netherlands. Its data collection was administered by the Criminal Justice Knowledge Center of the Netherlands Ministry of Security and Justice (WODC).

The present study aims to answer three research questions. First, we investigate whether time diaries are able to generate population estimates of offending. Second, we investigate whether time diaries are able to capture the development of individuals' involvement in offending over time. Third, we examine whether time diaries adequately capture the circumstances under which offending occurs.

2 Prior literature

2.1 Time use research on rare and short activities

2.1.1 Rare activities

Whether a time use instrument is able to capture rare activities depends on the *number of individuals* who engage in these activities and on *how often* they engage in these activities (Harvey, 1999). As only a limited amount of days can be covered in time diaries, most time diaries use relatively short reference periods. Respondents are asked to record their activities for one or two days, or at most for one whole week. However, activity patterns (or the frequency with which individuals engage in certain activities) may vary per season, per week, or even per day. Using a time diary with a short reference period is therefore not necessarily representative for longer periods of time. This is especially a problem for rare activities (Frazis and Stewart, 2012). As a potential solution – for rare activities specifically, Frazis and Stewart (2012) suggest that researchers collect data over multiple days for the same respondent. This allows for studying variance within and between individuals; having information about multiple days per person enables the comparison of activity patterns from one day with activity patterns from other days (for the same person). This gives at least some indication of the extent to which the activities on this day are representative for the individuals' daily routines. Nevertheless, this approach is not sufficient if the activities occur less often than the duration of the reference period. If you would question respondents about three days, and they are engaged in the activity on a weekly basis, you may still not capture the activity. More importantly, it would seem as if a respondent is not engaged in an activity *at all*, even though he or she is engaged in the activity on a weekly basis. Extending the reference period may increase the burden on respondents, thereby causing more non-response, and may decrease recall accuracy (Gershuny, 2012; Pentland et al., 1999). Scholars have stated that this limits the usefulness of time diaries for long-term population estimates (Frazis and Stewart, 2012; Gershuny, 2011).

Strategies other than increasing the reference period are for example to increase the sample size, or to select respondents who are known to frequently engage in the activity of interest (Kalton and Anderson, 1986). Alternatively, one could combine time diary data with stylized questionnaire data (habit-type items such as "How often do you engage in … activity?") to assess the distribution of activities across individuals and populations, as proposed by Gershuny (2012). He elaborates on a statistical technique developed for nutrition research by Tooze et al. (2006) and Kipnis et al. (2009). In this analytical framework, individual long-term time use estimates are calculated by multiplying respondents' daily participation probability with the predicted time engaged in the activity. These estimates are combined to estimate the distribution across populations (Gershuny, 2011; Gershuny, 2012). Gershuny (2012) showed that his method is applicable for a variety of infrequent activities, such as going to the cinema, watching sports, eating at a restaurant, and swimming.

2.1.2 Short activities

Time diaries with fixed time intervals (e.g., ten minutes, one hour) may underestimate the frequency of activities with a short duration. The larger the time interval, the more likely that multiple activities occur during the interval and, as longer activities are more likely to be reported, this leaves the shorter activities excluded (Ås, 1978). Previous studies indeed indicated that short activities were less likely to be captured by fixed interval time diaries than by open interval time diaries (Fleming and Spellerberg, 1999).

There are several ways to address this problem. A first option is to let the length of the fixed time interval be determined by the duration of the activity one intends to measure. This would mean that researchers who are interested in short activities need to define shorter time intervals (Fleming and Spellerberg, 1999). A disadvantage of using shorter fixed time intervals is the high costs of data collection. Also, using short fixed time intervals increases the burden on respondents, as the interview length is extended, and participants are required to recall briefer activities (Gershuny, 2012; Wikström et al., 2012a). It may also generate a lot of irrelevant data, such as on the "time spent making a cup of tea" (Wikström et al., 2012a, p. 74). A second option is the use summary questions, by for example asking at the end of every interview day: "During any part of the day did this activity take place?". The end of day summary question method was initially developed to capture secondary activities (Phipps and Vernon, 2009), but is applied in the Space-Time Budget method to measure activities of short duration that are of particular interest. The questions in the STB method are referred to as 'extra incidents' (Wikström, et al., 2010; Wikström, et al., 2012a). The use of end of day summary questions is relatively easy to implement and has been found to produce more realistic estimates of secondary activities (Phipps and Vernon, 2009; Schwartz, 2001). A third solution is to use time diaries with open time intervals. When using open intervals, respondents are asked to report the starting and ending time of each activity, regardless of the length of the activity. This strategy avoids issues with choosing the appropriate length of the intervals (Fleming and Spellerberg, 1999). Time diaries with open time intervals are for example applied in the

American Time Use Study (ATUS; Bureau of Labor Statistics, 2014). The Harmonized European Time Use Survey (HETUS) guidelines acknowledge the problem of underreporting short activities when using fixed time intervals, but nevertheless recommend applying fixed time intervals of ten minutes. Open interval time diaries are said to produce larger variation in data quality and to be more difficult to process (Eurostat, 2009).

2.2 Previous studies that applied time diaries to measure short and rare activities

Time diary data have been applied in earlier studies to capture short and rare activities. Most of the previous time diary studies on *rare* activities were concerned with identifying situational characteristics of those activities. These studies typically sampled individuals who were expected to frequently engage in the activities of interest. Margraf, Taylor, Ehlers, Roth, and Agras (1987), for example, collected systematic information on a large number of panic attacks to empirically establish a valid definition for such attacks. The study specifically selected patients who met the DSM criteria for panic disorder and used time diaries to assess symptom patterns, concurrent heart rate, physical activity, and the direct context of panic attacks. Abu-Arafeh and Callaghan (2004) applied time diaries to study the duration of migraine attacks among children and adolescents. All 720 participants attended a specialist headache clinic. Of the 231 participants who had migraine, 15 were asked to fill in prospective headache diaries to accurately record the duration of headache attacks. Epstein et al. (2009) applied time diaries to examine the craving for and use of cocaine and heroin, and included a volunteer sample of 114 cocaine- and heroin-abusing patients who were being treated with methadone. What these studies have in common is that they all used *non-random sampling* methods to deal with the infrequency of the target activity. Sturgis and Jackson (2003), on the other hand, used the *random sample* from the UK 2000 Time Use Survey to examine participation in sport- and cultural activities. They were interested in the association between social and individual characteristics and clusters of cultural activities. For that purpose, the time diary data turned out to be insufficient. They decided to rely on the questionnaire data, because many target activities (e.g., picking berries, mushrooms and herbs, fishing, doing gymnastics, going to art exhibitions) were recorded in less than one percent of the population. They concluded that "despite the undoubted superior quality of the time diary data in capturing the timing and other aspects of daily activities, the reference period of activity recording is too brief to warrant the sorts of analyses we wished to conduct" (Sturgis and Jackson, 2003, p. 6).

Previous time diary studies on *short (but frequent)* activities generally used open time intervals and exclusively asked about the activities of interest. Restroom visits, cigarette smoking, and social interactions are examples of relatively short, but frequent activities. Sampselle (2003), for example, used open interval time diaries and asked patients who experienced urinary incontinence to fill in for each day when they drank, how much, and which types of fluids, when they went to the toilet, and whether accidental leakage of urine had happened. Pa-

tients also had to specify the amount of urine leakage, whether they had the urge to urinate at that moment, and during which activity the accident took place. Surawy and Cox (1986) used an open interval time diary to examine the influence of mood and situation on smoking behavior. Respondents had to report their stress levels and the strength of their desire to smoke, prior to smoking their cigarette. After each cigarette, respondents were asked to report the context in which they smoked and whether they had enjoyed it. Dodge, Heimberg, Nyman and O'Brien (1987) investigated other-sex social interactions among high socially anxious and low socially anxious students by applying open interval time diaries. Respondents were asked to keep a diary for two weeks and to report the day, time, location, and duration of the social interaction, as well as the type of interaction and the nature of the relationship with the other person (among other things). Although many social interactions are short, they can of course vary tremendously in duration. In none of these studies it was clarified why an open time interval approach was chosen.

Offending may be even more difficult to capture than the activities previously discussed, because acts of offending are both rare *and* of short duration. In the following paragraph, we will discuss particular issues related to the measurement of offending in general, and to the measurement of offending with time diaries specifically.

2.3 Self-reported offending in stylized questionnaires and time diaries

Self-reports on offending are a valuable data source in criminological research, as they are perceived to be less biased than police reports (Thornberry and Krohn, 2000). A considerable percentage of crime goes unrecorded because the victims do not report to the police. The offenses that are not recorded by the police form the so called 'dark figure' of crime. Official crime records also reflect police prioritization policies. For example, variations in recorded traffic violations may also reflect variations in traffic surveillance.

Nevertheless, self-reports on offending, generally captured in stylized questionnaires, may also suffer from measurement problems and therefore not give a representative account of the actual crime rates. Sources of error exist, for example, because frequent offending adolescents may be less willing to cooperate or participate in studies on crime (Van San, 2008), and because the common practice of retrospectively questioning about a year may evoke recall problems: Studies have indicated that recall problems may arise when respondents are asked to report retrospectively about periods longer than a few months (Bachman and O'Malley, 1980; Clark, Fiebig and Gerdtham, 2008; Hill, 1985).

Despite these sources of error in measuring offending, there is a general consensus in criminology about the use of stylized questionnaires to capture self-reported offenses. In contrast, the time diary method is relatively new. Therefore, and because of the general concerns about the suitability of time diaries to capture short and rare activities, it is important to scrutinize the applicability of time diaries for measuring offending. We will do this by examining whether and how time diary estimates of offending differ from stylized questionnaire esti-

mates of offending. Comparisons between stylized questionnaire estimates and time diary estimates of other activities generally produce small correlations and suggest that stylized questionnaires overestimate the frequency and duration of activities (Juster, Ono and Stafford, 2003, Robinson, 1985, Stafford and Duncan, 1985). For example, studies comparing stylized questionnaire estimates with time diary estimates of household activities found systematic biases in the stylized questionnaire estimates (Kan, 2008, Kan and Pudney, 2008). Additionally, studies generally found that the average time spend on household activities are overestimated by the stylized questionnaires (Marini and Shelton, 1993; Press and Townsley, 1998 Schulz and Grunow, 2012). However, in the case of lengthy recall periods (i.e., six months, one year), the stylized questionnaire method may lead to *underreporting* of activities compared to time diaries (Hill, 1985), and to less accurate estimates in the case of irregular activities (Juster et al., 2003).

For the remainder of this study, it is important to keep in mind that offending differs from other behaviors in at least one aspect: Offending is likely to be more memorable, either because it is more exciting (Gudjonsson and Sigurdsson, 2007), or because of other reasons such as careful planning beforehand. We therefore expect that measurement error due to recall inaccuracy is less relevant for this particular behavior. From previous studies, we know that respondents have less trouble recalling activities that are important to them (Engle and Lumpkin, 1992) or that are more distinctive from other activities (Niemi, 1993). Habitual activities are more likely to be forgotten after a week or a month (Juster, 1985).

We now turn our focus to a time diary method that was specifically developed for criminological research: the Space-Time Budget method.

The Space-Time Budget method in criminology

The Space-Time Budget (STB) method is the first time diary method that has been applied on a large scale in criminological research. The method was developed by Wikström and colleagues and used in The Peterborough Adolescent and Young Adult Development Study (PADS+), conducted in England (Wikström et al., 2010; Wikström et al., 2012a). The STB method is designed to test the Situational Action Theory (Wikström, 2004; Wikström, 2005; Wikström, 2014). The STB method is derived from regular time diary methods, but extended with the geographical location of the activities and with a set of end of day summary questions about offending and other delinquent activities or events (Wikström et al., 2012a). The method has a reference period of four days (one Friday, one Saturday and the two most recent weekdays prior to the interview) and a fixed time interval of one hour, thus adding up to 96 hours per interview. The time diaries were administered retrospectively in one-on-one interviews by trained researchers, which differs from other time use studies that ask respondents to keep a diary of their activities throughout the day. For each hour, respondents were asked to report about their main activity, the people who were present, and the place where the activity took place. At the end of filling out each day, respondents were specifically asked about of-

fending ("Did you steal something or damaged something that belonged to someone else or were you involved in a fight?"), victimization ("Was something you possess stolen or broken? Has somebody beaten you, attacked you or did somebody start a fight with you?"), weapon carrying ("Did you carry a weapon at some point during this day?"), alcohol consumption, and drugs use ("Have you used alcohol or drugs during this day?"). The reported incidents in these end of day summary questions were added to the time diaries at the specific hour(s) that the incidents took place. Detailed overviews of the STB method as applied in PADS+ are given by Wikström et al. (2012a, 2012b), and as applied in SPAN by Hoeben et al. (2014).

Data derived with the Space-Time Budget method have been applied for several criminological research questions. Specifically, STB data have been applied to test the Situational Action Theory. This theory explains offending as the result of an interaction between a person's *crime propensity* and *criminogenic exposure* (Wikström, 2004; Wikström, 2005; Wikström, 2014). Individuals' crime propensity depends on the individual´s morality and ability to exercise self-control. Individuals' criminogenic exposure refers to the individuals´ exposure to settings with criminogenic characteristics. The STB method is unique in criminology in its ability to capture this 'criminogenic exposure': It enables capturing individual's activity fields. When combined with data on personal characteristics, STB data allow for examining whether exposure of individuals with certain characteristics (e.g., high crime propensity) to settings with certain characteristics (e.g., highly criminogenic settings) leads to higher crime rates.

STB data as collected in the PADS+ project have been mainly used to investigate the properties of the Situational Action Theory, but have also been applied to investigate, for example, the distances adolescents travel to crime locations and to study how individual activity patterns relate to demographic background variables (Wikström et al., 2012a). Situational aspects of offending have also been scrutinized in the PADS+ study, such as the influence of peer presence, lack of adult supervision, and unstructured activities (Wikström et al., 2012a). For more detailed overviews of studies conducted with the PADS+ STB data, see Wikström et al., (2010; 2012a)

In collaboration with the PADS+ staff, the Space-Time Budget method was translated and adjusted to the Dutch situation in the Study of Peers, Activities and Neighborhoods (SPAN). Data from the SPAN project have been used to examine, for example, situational causes of crime (Bernasco et al., 2013) and victimization (Averdijk and Bernasco, 2014), conditions under which 'time spent with peers' is related to adolescent offending (Weerman, Bernasco, Bruinsma and Pauwels, 2015a), whether the locations where adolescents 'hang out' are associated with offending (Hoeben and Weerman, 2014), and to investigate the extent to which parenting is related to time spent 'hanging out' with friends (Janssen, Deković, and Bruinsma, 2014).

As described in the previous sections, there have been some concerns about the suitability of time diaries for capturing short and infrequent activities. Rare activities may not be captured if

many respondents do not engage in the activities during the reference period. Short activities may be underreported if the fixed time interval is of longer duration than the activity. We will explore these potential limitations for offending: An activity that is both rare and of short duration. We are interested in whether (or to what extent) time diaries are able to generate offending population estimates and whether they are able to capture development of individuals' involvement in offending over time. Furthermore, we are interested in whether the end of day summary questions method (Phipps and Vernon, 2009; Schwartz, 2001) is useful for capturing the circumstances of the reported offenses. As such, we contribute to the time use literature by investigating the relevance of end of day summary questions for capturing short and rare activities. The end of day summary questions method has been developed, and thus far predominantly reviewed, for capturing secondary activities. In the Space-Time Budget method, the end of day summary questions approach was applied to overcome problems with the relatively long time intervals (of one hour) and to avoid underreporting of offenses. However, this approach may no longer adequately measure the conditions under which offending occurs, because the circumstances that were reported in a given hour are not necessarily those circumstances under which the offense occurred. A few studies have explored the validity of the Space-Time Budget method for measuring offending. We will briefly discuss their findings in the remainder of this section.

With regard to the Space-Time Budget methods' ability to generate *population estimates* of offending, Wikström et al. (2012a) found that high-frequent offenders have an increased chance to report offending during the four day-reference period of the time diary compared to low-frequent offenders. They matched the reported offending in the time diaries to the reported offending in the stylized questionnaires, completed by the same respondents. The respondents who reported at least one offense during the four time diary days, reported on average 110 crimes per year in the stylized questionnaire. Whereas the respondents who reported at least one offense (over the period of a year) in the stylized questionnaires, reported on average 32.7 crimes per year. This is of course to be expected, as adolescents who offend more often, are more likely to do so in any random four days. In fact, if we would assume that respondents who report offenses in the time diaries commit a crime every four days, they would commit roughly 92 crimes in one year, which is very close to the found average of reported offences in the questionnaires of 110 crimes per year. These findings suggest that a time diary with a brief reference period (of four days, or even less) may not sufficiently take into account the varying involvement in offending across individuals. Consequently, researchers who use time diaries to analyze rare activities, even if they are not interested in population figures, are likely to conduct their analyses for a sample that is not representative for the population.

With regard to the Space-Time Budget methods' ability to capture the *frequency of individual's involvement in offending*, Wikström and Butterworth (2006) found a significant correlation of 0.35 between offenses reported in time diaries and offenses reported by the same individuals in stylized questionnaires. With a follow-up dataset, Wikström et al. (2010; 2012a) found a significant correlation of 0.57 between the frequency of reported crimes in the time

diaries and the frequency of reported crimes by the same individuals in stylized questionnaires. Furthermore, they found that adolescents who reported crimes during the four days of the time diaries were more often registered in police records, compared to adolescents who had not reported crimes in the time diaries ($r = 0.20$, $p < .010$; Wikström et al., 2012a). These findings suggest that time diaries are at least to some extent able to capture individuals' involvement in offending.

Finally, with regard to the Space-Time Budget methods' ability to capture the *circumstances of offending*, Wikström et al. (2012a) found that offenses that were reported in the time diaries and crimes as registered by the police had similar temporal and spatial distributions. Conditions other than 'time' and 'space' may be relevant as well, such as the people who were present during the activity or whether the activity took place in a public setting. Although STB data have been used to examine these conditions (e.g. Wikström et al., 2012), they have not yet been validated in a comparison with external data. The present study will compare STB diary data with external data to examine whether time diaries adequately capture the circumstances under which offending occurs.

We are not aware of previous studies that investigated the ability of the Space-Time Budget method (or any other time diary method) to capture *development over time* of involvement in offending across individuals. Neither are we aware of studies that investigated the validity of the situational conditions of the offenses reported in time diaries, other than the spatial and temporal correlates investigated by Wikström et al. (2012a). The present study intends to fill this gap in the knowledge about the validity of time diaries to capture activities that are *both* short *and* rare.

In summary, the present study will examine whether time diaries are able to capture rare and short activities by addressing three research questions. First, we compare time diaries with stylized questionnaires (administered to the same sample) to test if population estimates of time diaries are able to take into account the varying involvement in offending across individuals. Second, we compare time diaries with stylized questionnaires (administered to the same sample) to investigate the hypothesized capability of time diaries to capture the development of individuals' involvement in offending over time. Third, we will compare the time diaries with external data to examine whether time diaries adequately capture the circumstances under which offending takes place.

3 Materials and method

3.1 Design

The present study examines whether the Space-Time Budget method is able to accurately capture acts of offending. To do so, STB data from the Study of Peers, Activities and Neighbor-

hoods (SPAN) were compared with data collected with two other instruments. First, the SPAN time diary data were compared with data from a stylized questionnaire that was administered among the same sample (also collected in the SPAN project). Second, the SPAN time diary data were compared with data from a stylized questionnaire that was developed to measure self-reported offending among the Youth Delinquency Survey (MZJ stylized questionnaire), a nationally representative survey administered by the Research and Documentation Centre (WODC) of the Netherlands Ministry of Security and Justice (Van der Laan and Blom, 2011). Data from the MZJ stylized questionnaires are largely representative for the Dutch adolescent population and include information about the situations in which offending occurs, which is particularly relevant for our research purposes.

3.2 Study of peers, activities and neighborhoods (SPAN)

The Study of Peers, Activities and Neighborhoods (SPAN) is a two-wave longitudinal study of the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), conducted amongst adolescents in the city and surrounding areas of The Hague (The Netherlands). The project incorporated several data collections, among which time diary interviews and stylized questionnaire interviews.

3.2.1 Sample

The sampling base of the SPAN study concerns Dutch youth between 11 and 20 years old who were registered at one of the ten participating secondary schools in the conurbation of The Hague (the Netherlands). For the data collection, 40 schools were approached, which represent approximately one third of the secondary schools in The Hague and surrounding regions. The city councilor of The Hague supported our study with a letter to the directors of the schools. As the municipality was conducting a study on health and welfare issues among over half of the secondary schools in The Hague, these schools were not approached for our study (Weerman, Bernasco, Bruinsma and Pauwels, 2015b). Of the 40 secondary schools, ten (25 percent) agreed to participate. Reasons for refraining from participation varied. Most of these schools were already participating in other research projects, others were afraid that the study would interfere with regular teaching or exams (Bernasco et al., 2013).

In total, 942 adolescents were selected to participate in the first wave of the data collection (all first and fourth graders of the participating schools), of which 843 adolescent completed the time diaries and the stylized questionnaires. Of the non-participating 99 adolescents, 35 completed only one of these instruments, three were ill during the data collection, fifteen were withdrawn from the study by their parents, 27 could not be reached, thirteen did not show up for their interviews, and six adolescents had moved to another school. In the second wave, 615 of the 843 adolescents participated again (73 percent). The main reasons for attrition were that adolescents simply refused participation, that they could not be contacted, that their parents refused to give permission, or that the adolescents repeatedly did not show up at ap-

pointments (Bernasco et al., 2013; Weerman et al., 2015a). Two adolescents in the second wave completed the time diaries, but were ill during the entire reference period (four days). They were excluded from the analyses. Thus, the analyses of the present study were conducted for 613 participants of the second wave. There was a time-span of approximately two years between the first wave (schoolyear 2008 - 2009) and the second wave of the data collection (schoolyear 2010 - 2011).

Participants of the study were between 11 and 18 years old (M = 14.14, SD = 1.70) in the first wave and between 13 and 20 years old (M = 16.02, SD = 1.69) in the second wave of the data collection. In both waves of the SPAN study, boys and girls were evenly represented (boys: 54.9 percent in wave one, 52.5 percent in wave two; girls: 45.1 percent in wave one, 47.5 percent in wave two). While most adolescents in the SPAN sample were native Dutch (55.1 percent in wave one; 55.7 percent in wave two), a relatively large part was from an ethnic minority background. Approximately eight percent was Moroccan, approximately nine percent Turkish, approximately seven percent Surinamese, approximately three percent Antillean, and approximately 18 percent from another ethnic minority background. Of the sample, twenty percent was following practical education, approximately 55 percent was following lower vocational education and a little over twenty percent was following medium or higher secondary education. Students following medium secondary education were underrepresented, compared to the Dutch adolescent population. All respondents lived within the city of The Hague or in nearby areas. The city is the third largest city in the Netherlands and densely populated.

3.2.2 Instruments

The time diary used in the SPAN project (the Space-Time Budget method) is a time diary with fixed time intervals of an hour and a reference period of four days. The time diaries are administered in one-on-one interviews. One-on-one interviews were preferred over handwritten diaries because they are more likely to result in complete records and because it allows interviewers to help respondents to recall activities and locations (Wikström et al., 2012a). For the Friday, Saturday and the two most recent weekdays prior to the interview, adolescents were asked, retrospectively, about their hourly activities, location, and who else was present in the given situation. Additionally, respondents were asked about offending, victimization, truancy, weapon carrying, alcohol and drug use. These questions were asked as summary questions, at the end of filling out a full day in the time diary. Answers on these end of day summary questions were immediately added to the diary data at the particular hour on which the summary question was applicable. A research-assistant (12 research-assistants in wave one; 15 research-assistants in wave two) administered the one-on-one interview, which took up about 45-50 minutes. To increase the ease and speed in which respondents' answers could be processed by the research-assistant, code lists were used that included codes for possible activities, locations and present people. If none of the existing codes represented the answer of the respondent, a new code was created by the research-assistant and added to the list.

Research suggests that the presence of an interviewer or other person during the data collection may enhance social desirable answering (e.g. Tourangeau and Yan, 2007). The risk of social desirable answering depends on whether respondents fear consequences of revealing information (Aquilino, Wright, and Supple, 2000); Clear instructions on confidentiality are found to have strong effects on the self-disclosure of participants (Woods and McNamara, 1980). To better capture offending, which is a sensitive topic, the interviews were conducted in a quiet area away from other participants. Interviewers explicitly informed participants about the anonymity of their responses in the diary and interviewers were instructed to refrain from any form of judgement. Additionally, participants were allowed to shake their head or nod in response to a question, instead of saying their answer out loud. In a further attempt to minimize bias in the reports, interviewers were selected that were slightly older than the respondents (i.e., interviewers had just graduated or were still in college). Studies have shown that both too much and too little social distance between interviewer and respondent would produce biasing effects (e.g. Dohrenwend, Colombotos and Dohrenwend, 1968; Nederhof 1985). Notwithstanding these efforts to minimize social desirable answering, the possibility that the offending measures are biased has to be taken into account when interpreting our findings. For more information on the Space-Time Budget method, see Wikström and Butterworth (2006) or Wikström et al. (2012a). The publications of Hoeben et al. (2014) and Wikström et al. (2012b) specifically discuss the practical application of the method in data collections.

Apart from the time diaries, SPAN respondents were asked to fill in a self-administered stylized questionnaire. This occurred also under the supervision of a research-assistant, for four respondents at the same time, during a time-span of 45-50 minutes. The questionnaire that was used in the SPAN project was based on the PADS+ questionnaire (Wikström and Butterworth, 2006; Wikström et al., 2012a) and intended to measure self-reported offending. The questionnaire included a variety of other variables as well, such as self-control, perceived parental monitoring, perceived neighborhood control, quality of the parent-child relationship and deviancy of peers. This background information was collected to complement the time diary data. To minimize social desirable answering, participants filled in the written questionnaires in silence – they were not allowed to interact with other participants – and we made sure the interviewers could not see the responses of the participants. When participants were finished, they handed in the sheets closed without their name on it.

3.3 The WODC Youth Delinquency Survey (MZJ) – National self-reports on offending

The Youth Delinquency Survey (MZJ stylized questionnaire), conducted by the criminal justice knowledge center of the Netherlands Ministry of Security and Justice (WODC) in 2010, is a cross-sectional study among adolescents. The questionnaire was designed to construct a clear picture of involvement in offending across Dutch adolescents and to enable the investi-

gation of trends in offending. The questionnaire also contains data on situational characteristics of offending. Such data is not available in the SPAN stylized questionnaire, which is why we added data generated by the MZJ stylized questionnaire in our study. The MZJ stylized questionnaire has been an important addition to Dutch police records of criminality (Van der Laan and Blom, 2011).

3.3.1 Sample

The sampling base of the MZJ study concerns Dutch youth between 10 and 17 years old who are legally residing in the Netherlands. The study applied a stratified sampling method. First, thirty municipalities in the Netherlands were randomly selected. Within each municipality, the addresses of adolescents were then also randomly selected. Adolescents were non-randomly selected on age (10-17 years old) and on ethnicity (Dutch, Moroccan, Turkish, Surinamese, Antillean and other ethnic backgrounds) which led to the selection of 4664 adolescents (Van der Laan and Blom, 2011). Ethnic minorities were non-randomly selected to create substantial group sizes. Previous research showed that group sizes became too small to compare groups, when ethnicity was randomly selected (Van der Laan and Blom, 2011). Of the selected adolescents, 4429 could be approached and 3030 of them eventually completed the questionnaire. Turkish and Moroccan adolescents were overrepresented in the non-participant group, but differences were small.

For the present study, we increased the comparability between the MZJ sample and the SPAN sample, by excluding the MZJ participants who followed primary education or who were younger than 11 years. It is important that both datasets have similar age ranges as research found that activity patterns change significantly in the years of adolescence (Wikström et al., 2012a). One participant had an extremely high score on one of the offending items and was also excluded. After this selection, 1849 of the 3030 MZJ respondents remained (61.0 percent). The participants were between 11 and 17 years old (M = 14.62, SD = 1.65) and the sexes were exactly evenly represented (boys 50.0 percent, girls 50.0 percent). The MZJ sample incorporated primarily adolescents with a native Dutch background (56.1 percent). Other participants had a Moroccan background (7.2 percent), Turkish background (8.5 percent), Surinamese (9.0 percent), Dutch Antillean (9.7 percent) or other ethnic background (9.5 percent). With regard to educational level, 1.5 percent of the MZJ participants were following practical education, 54.6 percent were following lower vocational education, and 43.9 percent were following medium or higher secondary education.

Comparisons with the SPAN participants from the first wave of the data collection showed that more of the MZJ participants were girls (χ^2 (1) = 5.56, p = .018, φ = .05), that the MZJ participants were older (t (2690) = -6.75, p < .001, CI of mean difference = -0.60 – -0.33, d = -0.28), and higher educated (χ^2 (2) = 328.19, p < .001, φ = .36). Comparisons with the SPAN participants from the second wave of the data collection showed that the MZJ participants

were significantly younger (t (2462) = 18.25, $p < .001$, CI of mean difference = 1.26 – 1.56, d = 0.84) and higher educated (χ^2 (2) = 263.09, $p < .001$, $\varphi = .34$).

3.3.2 Instruments

The MZJ stylized questionnaire was administered in a computer assisted one-on-one interview and intended to measure the prevalence of different types of youth offenses in the past year. To account for the sensitivity of questions on offending, the Computer Assisted Self-Interviewing method was used (CASI), in which no interviewer was present. The rest of the questions were administered by the interviewer using the Computer Assisted Personal Interviewing (CAPI) method. The questionnaire contained questions like "how many times have you offended in the previous twelve months?" and included several items on burglary, vandalism, violence, weapon possession and sex offenses (Van der Laan and Blom, 2011). In the present study, the sum of all the different types of offenses is used as one offending construct.

3.4 Comparing offending measures

Although all three measures of this study used comparable items to measure offending, they were not identical. Tables A1, A2, and A3 in the Appendix present overviews of the items and corresponding frequency distributions from, respectively, the SPAN time diaries, the SPAN stylized questionnaires, and the MZJ stylized questionnaires. The instruments also differed in their response categories. In the SPAN stylized questionnaires, offending was expressed in closed categories (e.g., 6 - 10 times, more than 10 times), whereas in the MZJ stylized questionnaires and SPAN time diaries, the responses for the offending items were open ended. The categories of the SPAN stylized questionnaires were recoded as following: 0 times = 0; 1 time = 1; 2 times = 2; 3 - 5 times = 4; 6 - 10 times = 8; more than 10 times = 10. Finally, the instruments differed in how the interviews were assessed. Whereas the SPAN time diaries were conducted as one-on-one interviews, the MZJ stylized questionnaires were assessed in Computer Assisted Personal Interviews (CAPI), and the SPAN stylized questionnaires were completed in written form under supervision of a research assistant.

To provide clarity about the comparability of offense measures across the different datasets, we included Table 1. Table 1 displays the descriptive statistics of the offending measures as reported in the SPAN time diaries, SPAN stylized questionnaires, and MZJ stylized questionnaires. The total number of offenses for the SPAN and MZJ questionnaires represents sum scores of the frequencies that were reported on each offending item. As the frequency categories of the SPAN stylized questionnaires were recoded (see previous paragraph), the offending rates derived from this instrument should be interpreted as approximate averages that are based on midpoints of categories rather than precisely reported frequencies.

Table 1
Descriptive statistics of offending in the SPAN and MZJ samples

	SPAN				MZJ
	TD T1	**TD T2**	**Quest. T1**	**Quest. T2**	
N (individuals)	843	613	843	613	1849
Not Offended	792 (94.0%)	588 (95.9%)	245 (29.1%)	226 (36.9%)	1295 (70.0%)
Offended	51 (6.0 %)	25 (4.1%)	598 (70.9%)	387 (63.1%)	554 (30.0%)
Total Nr. of offenses	69	32	7614	3834	4516
Mean	0.1	0.1	6.9	10.0	3.5
SD	0.4	0.3	13.2	18.7	21.5
Theft	5 (7.3%)	1 (3.1%)	2022 (26.6%)	1028 (26.8%)	909 (20.1%)
Vandalism	26 (37.7%)	14 (43.8%)	2350 (30.9%)	1133 (29.6%)	807 (17.9%)
Violence	35 (50.7%)	16 (50.0%)	1798 (23.6%)	853 (22.2%)	2159 (47.8%)
Other	3 (4.3%)	1 (3.1%)	1444 (19.0%)	820 (21.4%)	641 (14.2%)

The frequency of offending in the SPAN stylized questionnaires was expressed in categories. These were recoded as following: 0 times = 0; 1 time = 1; 2 times = 2; 3 - 5 times = 4; 6 - 10 times = 8; more than 10 times = 10. Abbreviations: TD = SPAN time diary; Quest = SPAN stylized questionnaire; T1 and T2 = first and second wave of data collection.
Source: Study of Peers, Activities and Neighborhoods (SPAN) and Youth Delinquency Survey (MZJ), own calculations.

Table 1 should be interpreted such that, for example, for the SPAN time diaries of wave one (first column), 843 adolescents completed the interview and 51 of them reported at least one offense. The other 729 adolescents did not report offenses during the four diary days. The total amount of reported offenses was 69, which is on average 0.1 offenses per person (standard deviation of 0.4) over the total sample of 843. Because the frequencies of offending in the time diaries in both waves were small, it was decided to combine the offending items into three categories: theft, violence, and vandalism. Of the 69 offenses that were reported, 7.3 percent concerned theft, 37.7 percent concerned vandalism, 50.7 percent concerned violence, and 4.3 percent concerned other types of offenses. The specific items per category are presented in Table A1 in the Appendix.

It is clear from Table 1 that the average number of reported offenses per individual vary across instruments. This is not surprising, given that the instruments cover different reference periods, with the SPAN time diary using a substantial smaller reference period than the other instruments (four days versus one year). Also, the proportions of offenses were somewhat different across instruments. Theft was relatively rare in the SPAN time diaries, which was not the case for the other instruments.

4 Results

As a first step in exploring whether time diaries offer an adequate way to measure offending, the SPAN time diary data were compared to the SPAN stylized questionnaire data. Both instruments were assessed among the same individuals living in the conurbation of The Hague (the Netherlands). In this first step, we investigated whether time diaries are able to sufficiently measure the distribution of offenses across individuals (whether the method is able to distinguish between non-offenders, medium-frequent offenders, and high-frequent offenders), and developments over time for the same individual (decreases and increases in delinquent involvement). The findings of these analyses are discussed in paragraphs 4.1 and 4.2.

Subsequently, we investigated whether the SPAN time diaries were able to capture situational circumstances of offending. In the SPAN time diaries, the end of the day summary question method was applied to measure offending. In these analyses, we compared the SPAN time diary data to stylized questionnaire data on offending collected by the criminal justice knowledge center of the Netherlands Ministry of Security and Justice (WODC); the MZJ stylized questionnaire. As the SPAN stylized questionnaires did not include situational information about offending it was not possible to examine this question within one sample. The findings of the analyses with the MZJ stylized questionnaire are discussed in paragraph 4.3.

4.1 Population estimates of adolescent offending

To compare the offending population estimates of the two data sets (SPAN time diary data and SPAN stylized questionnaire data), we estimated the average yearly offending rate per individual because, in the stylized questionnaires, we asked about offending *in the past year*. To do so, we had to conduct a weighting procedure for the SPAN time diary data. The SPAN time diaries were administered for four days: One Friday, one Saturday and two random weekdays. Activities on Fridays and Saturdays were therefore overrepresented compared to activities on the weekdays (Monday to Thursday), whereas activities on Sundays were not recorded. Because we lack information about activities on Sundays, we could not include them in the analyses. As for the overrepresentation of Fridays and Saturdays: These days make up fifty percent of all days administered in the time diaries, whereas they only take up 33.3 percent in an actual week (when Sunday is not included; two days out of six days makes 33.3 percent). We therefore assigned a weight of 0.67 (33.3/50.0) to the offenses that occurred on Fridays and Saturdays. The other days administered in the time diaries (Monday to Thursday) also make up fifty percent of all days administered in the time diaries, whereas they take up 66.7 percent of an actual week (in which Sunday is not included; four days out of six days makes 66.7 percent). We thus assigned a weight of 1.33 (66.7/50.0) to the offenses that occurred on Mondays to Thursdays.

A total of 101 offenses were reported during the SPAN time diary days in both waves, of which 42 offenses occurred on a weekday (Monday - Thursday) and 59 offenses occurred on

a Friday or Saturday. Given these numbers and the suggested weighting procedure, we expect that 95.39 offenses ((42*1.33) + (59*0.67)) would have been reported, had we asked our sample on four *random* days about their offending behavior. Given that a year has 365 days, we expect that 8704.34 offenses ((95.39/4 days)*365 days) would have been reported, had we asked our sample on every day of the year about their offending behavior. For one individual (of our combined sample of 1456 individuals), we would have then estimated an average of 5.98 offenses per year (8704.34/1456). Additionally, we have calculated the average offenses per year for an individual with an alternative approach that was suggested by Wikström et al. (2012a, p. 325). Whereas in the first approach the Sunday is treated as an average day (over both weekdays and weekend days), in the second approach the Sunday is treated as an average *weekday*. Based on the second approach, the estimated average number of offenses per year for one individual was 5.86[2].

In the SPAN stylized questionnaire, a total of 11448 offenses were reported across both waves of the data collection. These stylized questionnaires questioned about offending in the previous year. Therefore, we estimated an average of 7.86 offenses per year per individual (11448/1456). Based on these estimations of individual yearly offending rates, we conclude that individuals in our sample reported on average 31.4 to 34.1 percent more offenses in the SPAN stylized questionnaires (7.86 offenses per person per year) than in the SPAN time diaries (5.86 to 5.98 offenses per person per year, varying with calculation method). For a specification of the reported offenses in the SPAN time diaries and SPAN stylized questionnaires, see Tables A1 and A2 in the Appendix.

As a next step, we investigated whether time diaries are able to take into account the varying involvement in offending across individuals. To do so, we focused on all individuals who reported at least one offense in the SPAN time diary and on all individuals who reported at least one offense in the SPAN stylized questionnaire, and then compared their yearly offense rates based on the SPAN stylized questionnaires. In total, 76 individuals reported at least one offense during the four days of the SPAN time diaries and 986 individuals reported at least one offense during the year that was questioned in the SPAN stylized questionnaire. Of the 76 individuals who reported offenses in the SPAN time diary, five did not report any offense in the SPAN stylized questionnaire and the remaining 71 individuals reported an average of approximately 26.3 offenses per year[3] (*SD* = 25.3) in the questionnaires. This yearly offense rate

---

[2] In total, 59 crimes were reported in two weekend days. That is 59/2 = 29.5 crimes per weekend day. Next, 42 crimes were conducted across two weekdays, which is 42/2 = 21 crimes per weekday. To calculate the number of crimes per week, we have to multiply each outcome by the number of weekdays and weekend days that are present in a single week, respectively, and add them together. Thus: 29.5 crimes*2 weekend days + 21 crimes *5 weekdays = 164 crimes per week. If we multiply this by 52 (weeks per year), we find 8528 crimes per year. Divided by our sample of 1456, this leads to an estimate of 5.86 crimes per participant per year.

[3] In interpreting these averages, the reader should take into account that as the frequency categories of the SPAN stylized questionnaires were recoded, the offending rates derived from this instrument should be interpreted as approximate averages that are based on the summation of midpoints of categories rather than precisely reported frequencies.

of 26.3 offenses is higher than the average reported offense rate of the 986 adolescents who reported one or more offenses in the SPAN stylized questionnaire: They reported on average approximately 11.6 offenses per year (*SD* = 17.2). Thus, it seems that the offenses that are registered by the SPAN time diaries are more often committed by individuals who frequently engage in offending, compared to the offenses that are registered by the stylized questionnaires (See Table 1 for the yearly offending rates derived from the SPAN and MZJ questionnaires). These findings suggest that offenses committed by low-frequent offenders may not be sufficiently captured with time diaries.

4.2 Development of offending across individuals

To examine whether time diaries adequately capture decreases and increases in delinquent involvement for individuals over time, we computed 'difference scores' of offending between the two waves of the data collection. These waves had a time span of approximately two years in between. We did this for both instruments: The SPAN time diaries and the SPAN stylized questionnaires. For the SPAN time diaries, these difference scores reflected the difference in the total sum of incidents between wave one and wave two. In the SPAN stylized questionnaires, these difference scores reflected the difference in answer categories between wave one and wave two (0 = not engaged in offending in the past year; 1 = one time engaged in offending in the past year; 2 = two times; 3 = three to five times; 4 = six to ten times; 5 = more than 10 times engaged in offending in the past year). A value of zero on the difference score meant that the individual reported the same answer in both waves, a negative value indicated that the individual reported less involvement in offending in the second wave compared to the first wave, and a positive value indicated that the individual reported more involvement in offending in the second wave compared to the first wave. Descriptive statistics of the difference scores are shown in Table B in the Appendix.

Descriptive statistics of the offending measures in the SPAN stylized questionnaires and SPAN time diaries, are presented in Table 1, and Tables A1 and A2 in the Appendix. As the SPAN time diaries registered 'only' 101 offenses, we combined the reported incidents into three categories: Vandalism, theft, and violence. We made the same categories for the data from the SPAN stylized questionnaires. *Vandalism* incorporated the items on how often the adolescent had "damaged or destroyed something that not belonged to him or her". *Theft* incorporated the items on how often the adolescent had "stolen something (from a shop, a bike or scooter)". *Violence* incorporated the items on how often the adolescent had "beaten up somebody". Not many incidents of theft had been reported in the SPAN time diary. Therefore, we did not include theft in the analyses, but we included only violence and vandalism.

We then calculated Spearman correlations between the difference scores of both instruments. Spearman rank correlations are nonparametric test statistics that correct for positively skewed distributions. Significant relationships were found between the difference scores of both instruments for vandalism ($r = 0.14$, $p = .001$) and violence ($r = 0.19$, $p < .001$). This suggests

that reported violence and vandalism in both instruments (the SPAN time diary and the SPAN stylized questionnaire) showed somewhat similar developmental patterns for individuals over time. However, the correlations were small. This may indicate that one of the instruments is better in capturing developmental patterns of offending. We speculate that the longer reference period of the SPAN stylized questionnaires enables better measurement of individuals' development in engagement in offending over time, compared to the four-day reference period of the SPAN time diaries. Correlations were also examined for both waves of the data collection separately, in addition to the examination of the difference scores. The correlations for the independent waves were similar to the correlations found for the difference scores.

4.3 Capturing the circumstances of a short activity

As stated previously, the Space-Time Budget method included end of day summary questions to ask about offending (Phipps and Vernon, 2009; Schwartz, 2001). In these end of day summary questions, interviewers ask explicitly about the occurrence of (short) target activities at the end of each diary day about which participants were interviewed. They then ask during which (fixed) time intervals (e.g., in which hours), the activities occurred. In the Space-Time Budget diaries, respondents were asked at the end of each diary day whether they had been involved in offenses.

A potential disadvantage of the end of day summary question method is that it may not adequately measure the conditions under which target activities occur, because the circumstances that are reported in a given fixed time interval are not necessarily the circumstances of the activity. To study whether this method is useful for capturing the situational conditions of an offense, we compared the conditions of offenses reported in the SPAN time diaries with conditions of offenses reported in the Youth Delinquency Survey (MZJ stylized questionnaire). To do so, we isolated the offenders: The 76 adolescents who reported at least one offense in the SPAN time diary and the 554 adolescents who reported at least one offense in the MZJ stylized questionnaire. Descriptive statistics of these offenders and of the conditions under which their offenses were conducted are shown in Table C in the Appendix.

In the MZJ stylized questionnaire, individuals were asked about how often they had been involved in different types of offenses in the previous year. They were then asked about the situational characteristics of *their most recent* offense: Had it occurred in the presence of others, in a private or public location, during the week or in a weekend, at day or at night, and had they used alcohol prior to the offense? Individuals, who reported more than one *type* of offense, were asked about the situational conditions of each different type of offense. All situational conditions were recoded into dichotomous variables. The presence of others represented whether the respondent was alone during the offense (0), or whether he or she conducted the offense with others (1). The location of the offense was specified as either 'private' (0), when the participant was at home or at someone else's home, or 'public' (1), for offenses conducted elsewhere. Offenses that took place between Friday 6 P. M. and Sunday 6 P.M.

were coded as having occurred in the weekend (1), whereas offenses that took place on other moments were coded as having occurred during the week (0). Time of the offense was specified as day (0), for offenses that occurred between 6 A.M. and 6 P.M., and night (1), for offenses that occurred at other times. Alcohol use was coded as either having had alcohol (1), or not having had alcohol (0).

For the individuals who had reported more than one offense in the MZJ stylized questionnaire or SPAN time diaries, we recoded and dichotomized the situational conditions. For example, if a respondent would have reported three offenses in the SPAN time diary, of which one offense was conducted alone (score 0) and two were conducted in the presence of others (score 1), his or her mean score for 'presence of others' would be 0.67 and thus rounded to 1. This recoding allowed us to deal with dependency problems due to clustering of offenses within individuals.

We conducted several logistic regression analyses, predicting the odds that an offense had occurred under a specific condition. The results of these analyses are presented in Table 2.

Table 2
Logistic regressions predicting situational characteristics of offenses by sample (MZJ stylized questionnaire / SPAN time diary)

Variable	B	*SE*	*p*	*OR*	95 % CI
Alone/with others	1.19	0.35	<.001	3.27	[1.69, 6.32]
Private/public place	0.03	1.17	.949	1.03	[0.37, 2.88]
Week/weekend	0.28	0.27	.292	1.33	[0.78, 2.24]
Day/night	0.91	0.29	.001	2.50	[1.43, 4.35]
No alcohol/alcohol	0.55	0.45	.155	1.73	[0.81, 3.66]

Each row represents an independent logistic regression model, with the situational characteristics of offenses as the dependent variables. The coefficients in the table reflect the association between the sample variable (MZJ stylized questionnaire = 0; SPAN time diary =1) and the situational characteristic of the offense, controlled for offenders' gender, education and age. Coefficients for these control variables are not displayed. N = 76 for the SPAN time diary and N = 554 for the MZJ stylized questionnaire. The standard errors are bootstrapped on the basis of 10000 iterations. Abbreviations: *SE* = standard error; *OR* = odds ratio; CI = confidence interval.
Source: Study of Peers, Activities and Neighborhoods (SPAN) and Youth Delinquency Survey (MZJ), own calculations.

Each row of Table 2 represents a different logistic model predicting the odds that a reported offense had occurred under a specific situational condition. The main independent variable was the sample: The SPAN time diaries versus the MZJ stylized questionnaires. The models included controls for offenders' gender, education, and age. The first row in Table 2 represents a model that predicted the odds that the offense was conducted alone or with others. Results indicate that the offenses reported in the SPAN time diaries occurred more often in the presence of peers than the offenses reported in the MZJ stylized questionnaire (*OR* = 3.27,

$p < .001$). The third row in Table 2 represents a model that predicted the odds that offenses occurred during the night or during the day. Results show that the offenses reported in the SPAN time diaries occurred more often at night than the offenses reported in the MZJ stylized questionnaire ($OR = 2.50$, $p = .001$). The second, fourth and fifth row in Table 2 represent models that predict the odds that offenses occurred, respectively, in private or public places, during the week or weekend, and under the influence of alcohol. For none of these models significant differences were found between the SPAN time diaries and the MZJ stylized questionnaire.

We conclude that the situational conditions of reported offenses are for a large part similar in the two datasets (the SPAN time diary data and MZJ stylized questionnaire data). This suggests that the SPAN time diaries adequately captured most of the situational conditions under which offenses occur. It also suggests that the end of day summary question method is a valid solution to the problem of measuring short activities in time diaries, as it still enables the measurement of direct circumstances of those activities.

5 Discussion

The present study examined whether time diaries are suitable to measure short and rare activities, with a focus on one particular activity that is both short and rare; offending. We compared self-reported offending in a customized time diary to self-reported offending derived with two other instruments: Namely a stylized questionnaire that was administered among the same adolescent sample in the Dutch city of The Hague and a stylized questionnaire developed to measure self-reported offending in a nationally representative sample. The present study assessed whether time diaries were able to capture 1) long-term estimates of offending in the population, 2) changes in offending frequency over time across individuals, and 3) the circumstances under which offending occurred. To address these research questions, the study applied the Space-Time Budget method as developed by Wikström and colleagues (Wikström and Butterworth 2006; Wikström et al., 2010; Wikström et al., 2012a). Although this method was not developed to assess crime rates across time or the population, it used the end of day summary question method to capture offending (Phipps and Vernon, 2009; Wikström et al., 2012b), which makes it a unique instrument to assess validity questions regarding short and rare activities, particularly offending.

The findings indicated that time diaries may not fully capture all offenses. The population offense rate that was estimated based on the time diary data was lower than the population offense rate that was estimated based on stylized questionnaire data derived from the same sample. The offenses that were reported in the time diaries were relatively often committed by frequent offenders. These findings are consistent with previous findings from Wikström et al. (2012a). Researchers who use time diaries to analyze rare activities should be aware of the

possibility that their analyses are conducted with a sample that mainly represents people who frequently engage in the activity of interest.

The second research question of the present study concerned the development of individuals' involvement in offending over time. We found that the offenses that were reported in the time diaries showed similar patterns over time compared to the offenses that were reported in the stylized questionnaires. Although the correlations were small, this provides at least some support for the capability of time diaries to capture development over time of involvement in rare activities such as offending. Previous studies also reported small correlations between stylized questionnaires and time diaries for activities other than offending (e.g. Juster et al., 2003). Based on these results we cannot with certainty conclude which measure was most valid, as both methods apply self-reports about offending and we did not compare the diaries with administrative or observational measures. Even so, we cautiously suggest that researchers, who are solely interested in how activity patterns of rare activities change over time, may be better off using stylized questionnaire data. Stylized questionnaires generally have a longer reference period and are therefore more likely to capture the involvement in these activities by individuals who engage irregularly in these activities.

Finally, with regard to the third research question, we investigated whether the end of day summary question method is useful for capturing situational conditions of short activities such as offending. This method was initially developed to capture secondary activities (Phipps and Vernon, 2009), but was applied in the Space-Time Budget method to measure offending. We found that the end of day summary question method was indeed able to capture most of the situational conditions of the reported offenses, such as whether offenses had occurred in private or public places, during the week or in the weekend, or under the influence of alcohol.

In summary, our findings suggest that time diaries – which incorporate end of day summary questions about specific short and rare activities – might be useful for studying situational correlates of those activities (i.e. offending). Time diaries seem less useful for studying change in involvement in short and rare activities over time, or for estimating the prevalence of such activities across the general population.

The present study has some limitations that need to be addressed. First, to assess the validity of an instrument, it is best to compare it to another instrument that was administered from the same sample. This was not possible for our third research question, regarding the circumstances of offending, because the SPAN stylized questionnaire did not contain situational information about offenses. Therefore, the third research question was addressed by comparing estimates from the SPAN time diaries to estimates from the MZJ stylized questionnaires. It is possible that differences in measurement strategies and samples across both instruments confounded these comparisons. The reported situational conditions differed in two aspects: the offenses reported in the time diaries seemed to occur more often in the presence of peers (as opposed to alone) and more often at night (as opposed to during the day) than the offenses reported in the MZJ stylized questionnaires. The first difference, regarding the presence of

peers, may be explained by differing measurement strategies. The participants in the MZJ stylized questionnaire were explicitly asked about an offense and subsequently about whether peers were present. In the time diaries, the participants were first asked whether they were with peers in a specific hour, and were then asked whether an offense had occurred during that hour. The second difference, regarding daytime and nighttime offenses, may be explained by differences in the urban background of respondents. The sample from which the time diaries were administered (the SPAN sample) consisted of adolescents from a highly urban background: All participants lived within or nearby The Hague, which is the third largest city of the Netherlands. The sample from which the MZJ stylized questionnaires were administered (the MZJ sample), on the other hand, was representative for the Dutch adolescent population and therefore also included respondents from rural areas. It is possible that there is more nightly activity in urban areas compared to in rural areas, which may explain why the offenses reported in the time diaries seemed to occur more often at night. If this were the case, the findings reflect differences in measurement strategies and samples rather than validity problems of the time diary method.

A second limitation of the present study is that a small number of individuals reported an offense in the SPAN time diaries. Therefore, we could compare offending behavior of 76 offenders. With such small sample size, the risk of overlooking potential significant findings increases, as does the possibility of biased estimates. To deal at least partly with the issue of small sample sizes, we reported the bootstrapped standard errors to address the final research question regarding circumstances of offending.

A third limitation is that face-to-face interviews are potentially prone to social desirable answering, which may have affected our estimates for offending as captured with the SPAN time diaries. However, we have taken several precautions to ensure honest responses, as discussed in the method section. Also, offending incidents were still reported in these diaries, indicating that at least some offenders were willing to entrust us with such information.

A fourth limitation was that the response categories differed across the applied measurement instruments. The items in the SPAN stylized questionnaires were expressed in closed categories, whereas items in the MZJ stylized questionnaires and the SPAN time diaries were open ended. Relatedly, differences in interview style between interviewers may have resulted in differences between interviews and between measurement instruments. Both limitations, differences in response categories and interviewer effects, may have introduced measurement error.

To stimulate and inform future time use studies on offending or other short and rare activities, we have some suggestions about measuring such activities with time diaries. Our results suggest that the end of day summary question method, in which participants are asked specifically about the occurrence of a short activity, might be a valid solution for the problems concerned with measuring short activities. Other solutions – that have not been examined in the present study – for measuring short activities, are using small fixed time intervals or open

time intervals. With regard to measuring *rare* activities, we want to point again to a solution that has been widely adopted in previous time diary studies (e.g. Margraf et al., 1987; Epstein et al., 2009), namely that of including participants who are known to engage in the activity of interest. In the case of offending, candidate groups include prisoners or problem youth. This approach may not be ideal for researchers who are interested in the occurrence of a phenomenon across a general group of people, like we were interested in capturing offenses of 'normal' (not at-risk) youth. Nevertheless, this approach may become more relevant as the target activity is rarer. To illustrate this: Of the 1456 adolescents in our combined-waves sample, only 76 adolescents reported offenses in the SPAN time diary, which had a four-day reference period. Our findings indicate that the use of time diaries is especially restricted for estimating population rates of involvement in rare activities (population offense rates), and for estimating activity participation for individuals who do not often engage in the target activity (low-frequent offenders). Gershuny (2012) suggested that for relatively infrequent (much less than daily) activities researchers should combine their time diaries with habit-type items of questionnaires, such as: "How often do you engage in … activity?" (Gershuny, 2012). A final solution to better capture short and rare activities in time diaries, is to administer the interviews with smartphone time use apps. Time use smartphone apps can collect a lot of information about the activities and whereabouts of respondents, but still be less burdensome for respondents compared with time diaries that are administered in one-on-one interviews. Thus, with time use smartphone apps, one can more easily apply brief time intervals and long reference periods, which helps in capturing short and rare activities. Previous studies have already piloted smartphone time use apps (Sonck and Fernee, 2013).

Possibly many other solutions for measuring short and rare activities with time diaries are currently unexplored. Now criminology has started applying time use methods, the problems with capturing short and rare activities become increasingly relevant and are in need of a solution.

Appendix

Table A1
Frequencies of offenses in the SPAN time diaries
($N_{offenses}$ wave 1 = 69, $N_{offenses}$ wave 2 = 32)

Items	SPAN TD. T1 % (*N*)	SPAN TD. T2 % (*N*)
Theft		
Shoplift	4.3 (3)	0.0 (0)
From a person without violence	2.9 (2)	0.0 (0)
Home burglary	0.0 (0)	0.0 (0)
Burglary in a barn or garden	0.0 (0)	0.0 (0)
Burglary in a building	0.0 (0)	0.0 (0)
Breaking in a car to steal	0.0 (0)	0.0 (0)
Stealing a car	0.0 (0)	0.0 (0)
Stealing a scooter	0.0 (0)	0.0 (0)
Stealing a bike	0.0 (0)	0.0 (0)
Other theft	0.0 (0)	3.1 (1)
Vandalism		
Of a vehicle	0.0 (0)	3.1 (1)
Of a scooter	0.0 (0)	0.0 (0)
Of a bike	0.0 (0)	0.0 (0)
Of a house	0.0 (0)	3.1 (1)
Of a building	1.4 (1)	3.1 (1)
Of a lamppost or garage can	2.9 (2)	9.4 (3)
Graffiti	1.4 (1)	3.1 (1)
Other vandalism	31.9 (22)	21.9 (7)
Violence		
Hit or kicked someone	50.7 (35)	50.0 (16)
Other		
Theft from a person with violence	0.0 (0)	3.1 (1)
Traffic offenses	4.3 (3)	0.0 (0)

Abbreviations: SPAN TD. = SPAN time diary; T1 and T2 = wave 1 and wave 2. Numbers in this table represent the percentage of the total reported offenses. Numbers between brackets represent the total of reported offenses in the dataset.
Source: Study of Peers, Activities and Neighborhoods (SPAN), own calculations.

Table A2
Frequencies of offenses in the SPAN stylized questionnaires ($N_{offenses}$ wave 1 = 7614, $N_{offenses}$ wave 2 = 3834)

Items	SPAN Q. T1 % (*N*)		SPAN Q. T2 % (*N*)	
Theft				
Stolen something from a shop that was worth less than 5 euro	11.6	(885)	10.6	(406)
Stolen something from a shop that was worth more than 5 euro	2.6	(200)	5.5	(210)
Broken into a house to steal something		(85)	0.5	(21)
Broken into a car to steal something;	1.0	(76)	0.3	(13)
Broken in somewhere else to steal something	1.3	(98)	0.5	(18)
Robbed someone	1.0	(74)	1.2	(46)
Stolen anything from another person	2.2	(164)	2.9	(112)
Stolen a bike	3.5	(266)	3.3	(128)
Stolen a scooter	2.3	(174)	1.9	(74)
Vandalism				
Used graffiti or a marker on walls, doors or something else	15.2	(1157)	14.9	(570)
Damaged something not belonging to you	11.2	(853)	11.8	(454)
Set fire to something (for example in a building, a house, or a car)	4.5	(340)	2.8	(109)
Violence				
Beaten up a stranger on the streets	14.6	(1113)	14.2	(545)
Beaten up somebody which caused injuries for the person	9.0	(685)	8.0	(308)
Other				
Threatened someone to frighten the person or let the person do something for you	5.0	(380)	5.4	(208)
Sold weed or hash	4.5	(344)	5.2	(201)
Sold other drugs like XTC, cocaine, speed or something else	1.9	(143)	1.6	(64)
Bought something which you knew was stolen	5.1	(388)	6.7	(258)
Used a weapon	2.5	(189)	2.3	(89)

Abbreviations: SPAN Q. = SPAN stylized questionnaire; T1 and T2 = wave 1 and wave 2. Numbers in this table represent the percentage of the total reported offenses. Numbers between brackets represent the total of reported offenses in the dataset.
Source: Study of Peers, Activities and Neighborhoods (SPAN), own calculations.

Table A3
Frequencies of offenses in the
MZJ stylized questionnaires ($N_{offenses}$ = 4516)

Items	MZJ stylized questionnaire % (*N*)	
Theft		
Pickpocketed	1.1	(51)
Burgled	0.4	(18)
Stole something from a car	0.4	(19)
Stole a bike/scooter	3.1	(140)
Shoplifted less than 10 euro	12.4	(561)
Shoplifted more than 10 euro	2.7	(120)
Vandalism		
Damaged a vehicle	2.7	(120)
Damaged a residence	1.4	(63)
Damaged a bus, tram, metro or train	2.2	(98)
Damaged something else	11.6	(526)
Violence		
Beaten up somebody which caused injuries for the person	34.8	(1573)
Hit somebody without causing injuries	13.0	(586)
Other		
Used violence to steal something	0.1	(5)
Threatened someone with the intention to scare the other person	14.0	(630)
Threatened someone on the street to steal	0.1	(6)

Numbers in this table represent the total of reported offenses in the dataset. Absolute number of reported offenses with the percentages of the total reported offenses in brackets.
Source: Youth Delinquency Survey (MZJ), own calculations.

Table B
Descriptive statistics of the difference scores in the SPAN time diary and SPAN stylized questionnaire (*N* = 613)

	SPAN time diary					SPAN stylized questionnaire				
Variables	**Min**	**Max**	**Median**	**Mean**	**SD**	**Min**	**Max**	**Median**	**Mean**	**SD**
Vandalism	-3	2	0	0.00	0.26	-5	4	0	-0.14	0.97
Violence	-3	3	0	-0.01	0.32	-5	5	0	-0.19	1.14

N = 613, as 613 adolescents participated in both waves. Answer categories for the items from the SPAN stylized questionnaires were: zero times; 1 time; 2 times; 3-5 times; 6-10 times; more than 10 times. The numbers for the SPAN time diaries express the number of incidents in the four days of the time diaries. The difference scores express the differences in reported offending between wave one and wave two and can therefore be negative. A negative score indicates a decrease in reported offending over time and a positive score indicates an increase. Prior to calculating the difference scores, the missing values in the stylized questionnaires were imputed with the Expectation-Maximization method. For the SPAN time diaries, all abnormal days (on which the participant was ill, in the hospital, arrested, or on leave) were excluded. The remaining data was thought to better represent 'regular' days.
Abbreviations: Min. = minimum, Max. = maximum, *SD* = standard error.
Source: Study of Peers, Activities and Neighborhoods (SPAN), own calculations.

Table C
Characteristics of offenders and offenses reported in the time diary (*N* = 76 offenders) and MZJ stylized questionnaire (*N* = 554 offenders)

	SPAN time diary % (*N*)	MZJ stylized questionnaire % (*N*)
Characteristics of offenses[a]		
Present others		
Alone	15.8 (12)	40.3 (223)
With others	84.2 (64)	59.7 (331)
Place of offending		
Private	7.9 (6)	8.7 (42)
Public	92.1 (70)	91.3 (442)
Day of the week		
Weekday	51.3 (39)	62.3 (345)
Weekend	48.7 (37)	37.7 (209)
Time of offending		
Day	43.4 (33)	69.3 (384)
Night	56.6 (43)	30.7 (170)
Under influence of alcohol		
Yes	81.6 (62)	90.3 (500)
No	18.4 (14)	9.7 (54)
Characteristics of the offenders[b]		
Gender		
Boys	80.3 (61)	63.0 (349)
Girls	19.7 (15)	37.0 (205)
Educational level		
Practical education	11.0 (8)	1.7 (9)
Lower vocational education	74.0 (54)	59.1 (316)
Medium or higher secondary education	15.1 (11)	39.3 (210)

[a] Example of the 76 individuals who reported offenses in the SPAN time diaries, 12 (15.8 %) generally offended alone and 64 (84.2%) generally offended with others; [b] Offenders were between 12 and 20 years old (M = 14.92, *SD* = 1.98).

Source: Study of Peers, Activities and Neighborhoods (SPAN) and Youth Delinquency Survey (MZJ), own calculations.

References

Abu-Arafeh, I. and M. Callaghan (2004), Short migraine attacks of less than 2 h duration in children and adolescents, in: *Cephalalgia*, Vol. 24, No. 5, 333-338.

Aquilino, W.S., Wright, D.L. and A.J. Supple (2000), Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use, in: *Substance Use and Misuse*, Vol. 35, No. 6-8, 845-867.

Ås, D. (1978), Studies of time-use – Problems and prospects, in: *Acta Sociologica*, Vol. 21, No. 2, 125-141.

Averdijk, M. and W. Bernasco (2014), Testing the situational explanation of victimization among adolescents, in: *Journal of Research in Crime and Delinquency*, Vol. 52, No. 2, 151-180.

Bachman, J.G. and P.M. O'Malley (1980), When four months equal a year – An exploration of inconsistencies in students' monthly versus yearly reports of drug use – Monitoring the Future, Ann Arbor, Michigan, Institute for Social Research, The University of Michigan.

Bernasco, W., Ruiter, S., Bruinsma, G.J.N., Pauwels, L.J.R. and F.M. Weerman (2013), Situational causes of offending – A fixed-effects analysis of space-time budget data, in: *Criminology*, Vol. 51, No. 4, 895-926.

Buckle, A. and D.A. Farrington, (1984), An observational study of shoplifting, in: *British Journal of Criminology*, Vol. 24, No. 1, 63-73.

Bureau of Labor Statistics (2014), American time use survey user's guide – Understanding ATUS 2003 to 2013, Washington D. C., Bureau of Labor Statistics.

Clarke, P.M., Fiebig, D.G. and U.G. Gerdtham (2008), Optimal recall length in survey design, in: *Journal of Health Economics*, Vol. 27, No. 5, 1275-1284.

Dodge, C.S., Heimberg, R.G., Nyman, D. and G.T. O'Brien (1987), Daily heterosocial interactions of high and low socially anxious college students – A diary study, in: *Behavior Therapy*, Vol. 18, No. 1, 90-96.

Dohrenwend, B.S., Colombotos, J., and B.P. Dohrenwend (1968), Social distance and interviewer effects, in: *Public Opinion Quarterly*, Vol. 32, No. 3, 410-422.

Eurostat (2009), Harmonised European time use surveys – 2008 guidelines, Luxembourg, Office Official Publications of the European Communities.

Engle, P.L. and J.B. Lumpkin (1992), How accurate are time-use reports? – Effects of cognitive enhancement and cultural differences on recall accuracy, in: *Applied Cognitive Psychology*, Vol. 6, No. 2, 141-159.

Epstein, D.H., Willner-Reid, J., Vahabzadeh, M., Mezghanni, M., Lin, J. and K.L. Preston (2009), Real-time electronic diary reports of cue exposure and mood in the hours before cocaine and heroin craving and use, in: *Archives of General Psychiatry*, Vol. 66, No. 1, 88-94.

Fleming, R. and A. Spellerberg (1999), Using time-use data – A history of time-use surveys and uses of time-use data, Wellington, Statistics New Zealand.

Frazis, H. and J. Stewart (2012), How to think about time-use data – What inferences can we make about long- and short- run time use from rime-diaries?, in: *Annals of Economics and Statistics*, Vol. 105/106, 231-245.

Gershuny, J. (2011), Time-use surveys and the measurement of national well-being, Swansea, Office for national Statistics.

Gershuny, J. (2012), Too many zeros – A method for estimating long-term time-use from short diaries, in: *Annals of Economics and Statistics*, Vol. 105/106, 247-270.

Gudjonsson, G. H. and J. F. Sigurdsson (2007), Motivation for offending and personality – A study among young offenders on probation, in: *Personality and Individual Differences*, Vol. 42, No. 7, 1243-1253.

Harvey, A.S. (1999), Guidelines for time-use data collection and analysis, in: Pentland, W.E., Harvey, A.S., Lawton, M.P. and M.A. McColl (Eds.), *Time-use Research in the Social Sciences*, New York, Kluwer Academic Publishers Group, 19-45.

Hill, S.M. (1985), Patterns of time use, in: Juster, F.T. and F.P. Stafford (Eds.), *Time Goods and Well-Being*, Ann Arbor, University of Michigan, Institute for Social Research, 133-176

Hoeben, E.M., Bernasco, W., Weerman, F.M., Pauwels, L. and S. Van Halem (2014), The space-time budget method in criminological research, in: *Crime Science*, Vol. 3, No. 12, 1-15.

Hoeben, E.M. and F.M. Weerman (2014), Situational conditions and adolescent offending – Does the impact of unstructured socializing depend on its location?, in: *European Journal of Criminology*, Vol. 11, No. 4, 481-499.

Janssen, H.J., Deković, M., and G.J.N. Bruinsma (2014), Parenting and time adolescents spend in criminogenic settings – A between- and within-person analysis, in: *British Journal of Criminology*, Vol. 54, No. 4, 551-567.

Juster, F.T., Ono, H. and F.P. Stafford (2003), An assessment of alternative measures of time use, in: *Sociological Methodology*,Vol. 33, No. 1, 19-54.

Juster, F.T. (1985), Conceptual and methodological issues involved in the measurement of time-use, in: Juster, F.T. and F.P. Stafford (Eds.), *Time Goods and Well-Being*, Ann Arbor, University of Michigan, Institute for Social Research, 19-28.

Kan, M.Y. (2008), Measuring housework participation – The gap between ''stylised'' questionnaire estimates and diary-based estimates, in: *Social Indicators Research*,Vol. 86, No. 3, 381-400.

Kan, M.Y. and S. Pudney (2008). Measurement errors in stylised and diary data on time use, in: *Sociological Methodology*, Vol. 38, No. 1, 101-132.

Kalton, G. and D.W. Anderson (1986), Sampling rare populations, in: *Journal of the Royal Statistical Society*, Vol. 149, No. 1, 62-82.

Kipnis, V., Midthune, D., Buckman, D.W., Dodd, K.W., Guenther, P.M., Krebs-Smith, S.M., Subar, A.F., Tooze, J.A., Carroll, R.J. and L.S. Freedman (2009), Modeling data with excess zeros and measurement error – Application to evaluating relationships between episodically consumed foods and health outcomes, in: *Biometrics*, Vol. 65, No. 4, 1003-1010.

Margraf, J., Taylor, B., Ehlers, A., Roth, W.T. and W.S. Agras (1987), Panic attacks in the natural environment, in: *Journal of Nervous and Mental Disease*, Vol. 175, No. 9, 558-565.

Marini, M.M. and B.A. Shelton (1993), Measuring household work – Recent experience in the United States, in: *Social Science Research*, Vol. 22, No. 4, 361-382.

Nederhof, A.J. (1985), Methods of coping with social desirability bias – A review, in: *European Journal of Social Psychology*, Vol. 15, No. 3, 263-280.

Niemi, I. (1993), Systematic error in behavioural measurement – Comparing results from interview and time budget studies, in: *Social Indicators Research*, Vol. 30, No. 3, 229-244.

Osgood, D.W., Wilson, J.K., O'Malley, P.M., Bachman, J.G. and L.D. Johnston (1996), Routine activities and individual deviant behavior, in: *American Sociological Review*, Vol. 61, No. 4, 635-655.

Pentland, W.E., Harvey, A.S., Lawton, M.P. and M.A. McColl (1999), Time-use research in the social sciences, New York, Kluwer Academic Publishers Group.

Phipps, P.A. and M.K. Vernon (2009), Twenty-four hours – An overview of the recall diary method and data quality in the American Time-use Survey, in: Beli, R.F., Stafford F.P. and D.F. Alwin (Eds.), *Calendar and time-diary – Methods in life course research*, London, Sage Publications, 109-129.

Press, J.E. and E. Townsley (1998), Wives' and husband's housework reporting – Gender, class, and social desirability, in: *Gender and Society*, Vol. 12, No. 2, 118-218.

Robinson, J.P. (1985), The validity and reliability of diaries versus alternative time use measures, in: Juster, F.T. and F.P. Stafford (Eds.), *Time Goods and Well-Being*, Ann Arbor, University of Michigan, Institute for Social Research, 19-28.

Sampselle, C.M. (2003), Teaching women to use a voiding diary – What 'mind over bladder' can accomplish, in: *The American Journal of Nursing*, Vol. 103, No. 11, 62-64.

Schulz, F. and D. Grunow (2012), Comparing diary and survey estimates on time use, in: *European Sociological Review*, Vol. 28, No. 5, 622-632.

Schwartz, L.K. (2001), Minding the children – Understanding how recall and conceptual interpretations influence responses to a time-use summary question, Washington D.C., Bureau of Labor Statistics.

Sonck, N. and H. Fernee (2013), Using smartphones in survey research – A multifunctional tool, The Hague: The Netherlands Institute for Social Research.

Stafford, F.P. and G.J. Duncan (1985), The use of time and technology in the United States, in: Juster, F.T. and F.P. Stafford (Eds.), *Time Goods and Well-Being*, Ann Arbor, University of Michigan, Institute for Social Research, 245-282.

Sturgis, P. and J. Jackson (2003), Examining participation in sporting and cultural activities – Analysis of the UK 2000 Time Use Survey phase 2, London, Department for Culture, Media and Sport.

Surawy, C. and T. Cox (1986), Smoking under natural conditions – A diary study, in: *Personality and Individual Differences*, Vol. 8, No. 1, 33-41.

Thornberry, T.P. and M.D. Krohn (2000), The self-report method for measuring delinquency and crime, in: *Criminal Justice*, Vol. 4, No. 1, 33-83.

Tooze, J.A., Midthune, D., Dodd, K.W., Freedman, L.S., Krebs-Smith, S.M., Subar, A.F., Guenther, P.M., Carroll, R.J. and V. Kipnis (2006), A new method for estimating the usual intake of episodically-consumed foods with application to their distribution, in: *Journal of the American Dietetic Association*, Vol. 106, No. 10, 1575-1587.

Tourangeau, R and T. Yan (2007), Sensitive questions in surveys, in: *Psychological Bulletin*, Vol. 133, No. 5, 859-883.

Van der Laan, A.M. and M. Blom (2006), Jeugddelinquentie – Risico's en bescherming – Bevindingen uit de WODC monitor zelfgerapporteerde jeugdcriminaliteit 2005, Den Haag, Boom Juridische uitgevers/WODC.

Van der Laan, A.M. and M. Blom (2011), Jeugdcriminaliteit in de periode 1996-2010 – Ontwikkelingen in zelfgerapporteerde daders, door de politie aangehouden verdachten en strafrechtelijke daders op basis van de monitor jeugdcriminaliteit 2010, Den Haag, Boom Juridische uitgevers/WODC.

Van San, M. (2008), Een wereld van verschil – Over het onderzoek naar criminaliteit en etniciteit in Beligië en Nederland, in: Weijers, I. and C. Eliaerts (Eds.), *Jeugdcriminologie – Achtergronden van jeugdcriminaliteit*, Den Haag, Boom Juridische Uitgevers, 283-300.

Weerman, F.M., Bernasco, W., Bruinsma, G.J.N. and L.J.R. Pauwels (2015a), When is spending time with peers related to delinquency? – The importance of where, what, and with whom, in: *Crime and Delinquency*, Vol. 61, No. 10, 1386-1413.

Weerman, F.M., Bernasco, W., Bruinsma, G.J.N. and L.J. Pauwels (2015b), Gender differences in delinquency and situational action theory – A partial test, in: *Justice Quarterly*, Vol. 33, No. 7, 1128-1209.

Wikström, P.-O.H. (2004), Crime as alternative – Towards a cross-level situational action theory of crime causation, in: McCord, J. (Ed.), *Beyond empiricism – Institutions and intentions in the study of crime*, New Brunswick, Transaction, 1-37.

Wikström, P.-O.H. (2005), The social origins of pathways in crime – Towards a developmental ecological action theory of crime involvement and its changes, in: Farrington, D. (Ed.), *Integrated developmental and life-course theories of offending*, New Brunswick, Transaction, 211-245.

Wikström, P.-O.H. (2014), Why crime happens – A situational action theory, in: G. Manzo (Ed.), *Analytical sociology – Actions and networks*, Hoboken, John Wiley and Sons Ltd, 71-94.

Wikström, P.-O.H. and D.A. Butterworth (2006), Adolescent crime – Individual differences and lifestyles, Cullumpton, Willan Publishing.

Wikström, P.-O.H., Ceccato, V., Hardie, B. and K. Treiber (2010), Activity fields and the dynamics of crime – Advancing knowledge about the role of the environment in crime causation, in: *Journal of Quantitative Criminology*, Vol. 26, No. 1, 55-87.

Wikström, P.-O.H., Oberwittler, D., Treiber, K. and B. Hardie (2012a), Breaking rules – The social and situational dynamics of young people's urban crime, Oxford, Oxford University Press.

Wikström, P.-O.H., Treiber, K. and B. Hardie (2012b), Examining the role of the environment in crime causation: Small-area community surveys and space-time budgets, in: Gadd, D., Karstedt, S. and S.F. Messner (Eds.), *The Sage handbook of criminological research methods*, Los Angeles, Sage Publications, 111-127.

Woods, K.M. and J.R. McNamara (1980), Confidentiality – Its effect on interviewee behavior, in: *Professional Psychology*, Vol. 11, No. 5, 714-721.