

## Dynamic modelling of n-of-1 data: powerful and flexible data analytics applied to individualised studies

Rute Vieira, Suzanne McDonald, Vera Araújo-Soares, Falko F. Sniehotta & Robin Henderson

To cite this article: Rute Vieira, Suzanne McDonald, Vera Araújo-Soares, Falko F. Sniehotta & Robin Henderson (2017) Dynamic modelling of n-of-1 data: powerful and flexible data analytics applied to individualised studies, *Health Psychology Review*, 11:3, 222-234, DOI: [10.1080/17437199.2017.1343680](https://doi.org/10.1080/17437199.2017.1343680)

To link to this article: <https://doi.org/10.1080/17437199.2017.1343680>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Accepted author version posted online: 20 Jun 2017.  
Published online: 06 Jul 2017.



[Submit your article to this journal](#)



Article views: 940



[View related articles](#)



[View Crossmark data](#)



Citing articles: 3 [View citing articles](#)



## Dynamic modelling of n-of-1 data: powerful and flexible data analytics applied to individualised studies

Rute Vieira<sup>a</sup>, Suzanne McDonald<sup>a</sup>, Vera Araújo-Soares<sup>a</sup>, Falko F. Sniehotta<sup>b</sup> and Robin Henderson<sup>c</sup>

<sup>a</sup>Institute of Health & Society, Newcastle University, Newcastle upon Tyne, UK; <sup>b</sup>Fuse, UKCRC Centre for Translational Research in Public Health, Institute of Health & Society, Newcastle University, Newcastle upon Tyne, UK; <sup>c</sup>School of Mathematics and Statistics, Newcastle University, Newcastle upon Tyne, UK

### ABSTRACT

N-of-1 studies are based on repeated observations within an individual or unit over time and are acknowledged as an important research method for generating scientific evidence about the health or behaviour of an individual. Statistical analyses of n-of-1 data require accurate modelling of the outcome while accounting for its distribution, time-related trend and error structures (e.g., autocorrelation) as well as reporting readily usable contextualised effect sizes for decision-making. A number of statistical approaches have been documented but no consensus exists on which method is most appropriate for which type of n-of-1 design. We discuss the statistical considerations for analysing n-of-1 studies and briefly review some currently used methodologies. We describe dynamic regression modelling as a flexible and powerful approach, adaptable to different types of outcomes and capable of dealing with the different challenges inherent to n-of-1 statistical modelling. Dynamic modelling borrows ideas from longitudinal and event history methodologies which explicitly incorporate the role of time and the influence of past on future. We also present an illustrative example of the use of dynamic regression on monitoring physical activity during the retirement transition. Dynamic modelling has the potential to expand researchers' access to robust and user-friendly statistical methods for individualised studies.

### ARTICLE HISTORY

Received 26 February 2017  
Accepted 24 May 2017

### KEYWORDS

N-of-1 methods; dynamic regression; single-case; individualised study

## Introduction

Between-participant studies, such as cohort studies or randomised controlled trials (RCTs), are the most prevalent research study designs, whether the aim is to identify predictors of response or estimate the population-level effect of an intervention. However, the evidence generated often has limited applicability to individuals seen in every-day practice (Rothwell, 2005) as it usually aims to estimate the average effect of specific predictors for a given population (Duan, Kravitz, & Schmid, 2013).

Furthermore, there is a growing interest in the opportunities individualised quantitative approaches can offer when developing or testing an intervention. Practices aimed at personalising the care of an individual have been attracting considerable attention in recent years (Lillie et al., 2011; Person Centred Care/Coalition for Collaborative Care/Medical directorate, 2016).

N-of-1 studies (see Box 1) are recognised as a valid and efficient mechanism to inform intervention development (Lillie et al., 2011), evaluate individualised evidence-based interventions (Craig et al.,

**CONTACT** Rute Vieira ✉ [rute.vieira@newcastle.ac.uk](mailto:rute.vieira@newcastle.ac.uk); [rute.vieira@ncl.ac.uk](mailto:rute.vieira@ncl.ac.uk)

Supplemental data for this article can be accessed at <https://doi.org/10.1080/17437199.2017.1343680>

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2008) and test theory (Johnston & Johnston, 2013; Naughton & Johnston, 2014). A renewed interest in n-of-1 studies has arisen with the rapid development of electronic health information technology (e.g., fitness trackers) within different areas of clinical care (Barr et al., 2015) and health behaviour research (McDonald et al., 2017; O'Brien, Philpott-Morgan, & Dixon, 2016). The ability to design studies to test hypotheses and interventions at the individual level is also key for studying rare conditions, where between-participant studies are not feasible due to the small numbers of affected individuals (e.g., in xeroderma pigmentosum, a rare autosomal recessive genetic disorder; Sainsbury et al., 2016). Moreover, n-of-1 studies provide the scientific basis to tailor interventions to individuals in a way that no other set of research designs can achieve (Sniehotta, Presseau, Hobbs, & Araújo-Soares, 2012). This process can be easily integrated into mobile applications and digital interventions which track the user's behaviour over time (cf. McDonald et al., 2017 for a review of n-of-1 applications in health behaviour research).

Box 1. Important facts about n-of-1 studies.

- (1) **It is an individualised study** and its main aim is not to infer population-level parameters but to reach conclusions for the individual under study.
- (2) **It is always a time-series** as the only available way to measure the variability within an individual is by obtaining repeated measurements over time. Likewise, the potential predictors need to vary over time.
- (3) **The number of repeated measurements is the sample size of an n-of-1 study.** More measurements lead to a better measure of the variability in the outcome of interest and improved precision of parameter estimates.
- (4) **'Bespoke' design** → **'Bespoke' analysis.** As these are individualised studies, each study is designed according to real-world considerations which are specific to each individual. How, when and which variables are being collected depends on the individual circumstances. N-of-1 designs may vary substantially and reflect greater creativity than other, more commonly used designs. The statistical analysis plan will have to mirror the design.
- (5) **Aggregated analysis of n-of-1 data is a possibility**, allowing different questions to be answered about generalisability across people. Random-effects meta-analysis, mixed models and the summary measures approach are the most commonly used statistical methods for aggregated n-of-1 data (Araújo, Julious, & Senn, 2016).

## Many terminologies – one data structure

Many different terminologies are currently used to describe n-of-1 studies (e.g., single-case, single-subject and single-patient studies). While all terminologies are acceptable, there is inconsistency in their use across published statistical methodologies and methodological reviews (e.g., Duan et al., 2013; Perdices & Tate, 2009; Shadish, 2014). Therefore, it is important to clarify this interchangeability in terminology when discussing the statistical aspects and state-of-the-art statistical methods for n-of-1 studies.

In medicine, the term n-of-1 mainly refers to an interventional design used to test medical or pharmacological treatments (Barr et al., 2015; Duan et al., 2013; Kravitz, Duan, & the DEClDE Methods Center N-of-1 Guidance Panel, 2014), while in psychology, it often also refers to a range of designs including observational studies, multiple baseline, alternating treatments, changing criterion and pre–post (AB) designs, which are commonly used to identify predictors of behaviour (Barlow, Nock, & Hersen, 2008; Kazdin, 2011; Kwasnicka, Dombrowski, White, & Sniehotta, 2017; McDonald et al., 2017; O'Brien et al., 2016; Quinn, Johnston, & Johnston, 2013; Shamsseer et al., 2015). More recently, n-of-1 trials have been named micro-randomised trials when specifically applied to data obtained from activity trackers and mobile phones (Dempsey, Liao, Klasnja, Nahum-Shani, & Murphy, 2015; Klasnja et al., 2015; Law, Edirisinghe, & Wason, 2016; Liao, Klasnja, Tewari, & Murphy, 2016). Terms which describe the nature of the data itself have also been used to describe n-of-1 studies. For example, Borckardt et al. (2008) propose a simulation method for the analysis of 'case-based time-series design' in psychotherapy, which has been applied in the analysis of n-of-1 data (O'Brien et al., 2016), while Velicer and Fava (2003) refer simply to 'time-series analysis' when analysing repeated observations on a single individual at regular intervals.

Independently of the terminology used, the data that results from n-of-1 studies have a time-series structure resulting from measuring one or several variables over time in a given unit, which is often an individual, perhaps measured in a daily context. This type of design and the shape and structure of the data have been commonly used in other fields. For example, time-series designs have been employed in econometrics to study market behaviour and economic forecasting, in engineering for evaluating quality control and in the analysis of political processes (Box & Jenkins, 1970; Glass, Willson, & Gottman, 1975; Pevehouse & Brozek, 2008). Across these applications, the objectives of such longitudinal designs are usually to identify predictors of response, describe adaptive changes over time or predict future outcomes given prior history.

### Taking time seriously

Time-series data have a natural temporal ordering and are typically not independent as the same individual is repeatedly measured, thus the data may exhibit some form of serial dependence or autocorrelation (i.e., the value of one observation depends, at least partly, on the value of one or more of the preceding observations in the series). When ignoring autocorrelation within time-series data, the standard errors are likely to be underestimated or overestimated, depending on the presence of positive or negative autocorrelation. The former increases the risk of Type I error (i.e., identifying a non-existent effect), while the latter increases the risk of Type II error (i.e., not identifying a true effect). Either way, it results in unreliable measures of statistical significance. Therefore, it is essential to acknowledge that n-of-1 data have a time-series structure and that autocorrelation, if present, needs to be accounted for when considering statistical analysis.

There is currently a lack of consensus concerning the most effective analysis methods for behavioural and psychological n-of-1 methods (Shamseer et al., 2015). Visual analysis of the slope, variability and patterns within the data has been used preferentially for time-series with low numbers of observations (Cooper, Heron, & Heward, 2007). However, it does not fully utilise the potential of contemporary health research methodology as more repeated measures are collected. Also, the statistical determination of effect sizes (and their precision) has proven to be an invaluable aid in decision-making and future research (e.g., sample size calculation, meta-analysis; Shamseer et al., 2015). But despite the apparent simplicity of an n-of-1 study design, the majority of previous n-of-1 studies were not statistically analysed. This may be related to the perceived complexity and uncertainty of the required statistical analysis (Kravitz et al., 2009; Shadish, 2014). Gabler, Duan, Vohra, and Kravitz (2011) reviewed 108 single-case trials reported in the medical literature during 1985–2010 and found that around 50% used a statistical approach to determine a superior treatment. However, in a recent systematic review on the use of n-of-1 methods in health behaviour research (McDonald et al., 2017), only 25% of the 39 studies included in the review used statistical approaches for n-of-1 data analysis. Although not entirely clear why statistical methods are more commonly used in clinical research, n-of-1 studies in this area usually use the RCT design with the aim to compare treatments and test hypotheses, which often implies the use of statistical methodologies. In addition, behaviour research has used a variety of n-of-1 designs, which is in part related to the irreversibility of some interventions (e.g., information provision). Only more recently, behaviour research has been using n-of-1 RCT to test reversible interventions components (those which can be removed; e.g., Sniehotta et al., 2012). Differences in the designs used in both fields might therefore play some role in explaining the difference in prevalence in the use of statistical approaches as more complex designs might hinder the use of statistical methods. However, in both fields, there was a tendency to use simpler and, arguably, less appropriate statistical methods.

Several statistical approaches have been proposed for determining effect sizes in n-of-1 studies: from simple paired *t*-tests and standardised effect sizes (e.g., *d* statistics, odds ratio and correlation coefficients), double bootstrap methods and the Cochrane–Orcutt approach (Bagian, King, Mills, & McKnight, 2011; McKnight, McKean, & Huitema, 2000; Naughton & Johnston, 2014) to more

complex time-series methodologies like ARMA and ARIMA modelling (Mills, 1990), simulation modelling analysis (Borckardt et al., 2008; Nash, Borckardt, Abbasa, & Gray, 2011), Bayesian statistics (Swaminathan, Rogers, & Horner, 2014) or adaptive treatment regimens methods in which the treatment decision at each step is made sequentially by utilising the cumulative data previously collected in the trial (Bembom & van der Laan, 2008; Cheung, Chakraborty, & Davidson, 2015; Henderson, Ansell, & Alshibani, 2010; Mao & Cheung, 2016). As for the simpler approaches, conventional parametric and non-parametric statistics often assume that observations are independent. Therefore, most of these techniques ignore the presence of autocorrelation and are usually not appropriate to analyse n-of-1 data (Parker, Vannest, & Davis, 2011; Shadish, 2014). A much more familiar and flexible approach is the use of regression-based models as they can account for autocorrelation while assuming different distributions for the outcome (e.g., count, categorical and non-normal continuous outcomes). Additional advantages are the possibility of testing for a non-linear time trend as an integral part of the model-selection process. Although non-parametric tests of non-linearity in time-series exist, including the Hinich test (Hinich, 1982) and the bispectral test (Rusticelli, Ashley, Dagum, & Patterson, 2008), which could be potentially used in addition to non-parametric or more simple parametric approaches, there is no evidence of their use when analysing n-of-1 data. Nevertheless, the added complexity of some of the high-end time-series regression-based methods may limit feasibility due to the high level of statistical expertise required for their use. There is a pressing need not only to facilitate access to these sophisticated and appropriate methodologies for n-of-1 analysis, but also to other suitable approaches which have been commonly used in other fields to analyse this type of data.

In this article it is not our intention to provide a critical review of current statistical approaches used to analyse n-of-1 data. Instead we aim to help general understanding about the statistical analysis of n-of-1 studies and strengthen the quality of evidence generated in future studies by introducing and promoting the use of dynamic models, also referred to as autoregressive distributed lag models (Hendry, 1995), in behavioural research.

### What is dynamic modelling?

Dynamic modelling accounts for the effect of past on the future by including lagged variables representing the past history of the predictors and outcome (which adjusts for the presence of autocorrelation) in an otherwise conventional multiple regression model. Focusing on how dynamic modelling deals with autocorrelation, Keele and Kelly (2006) used Monte Carlo analysis to compare the performance of dynamic models with several other time-series models including ARMA, Cochrane–Orcutt and Prais–Winsten (1954) regressions, which are the most likely alternatives for dealing with autocorrelation. The analysis showed that if the process is dynamic, i.e., if there is an effect of the past on the current values of the process being studied, then the estimates provided by the dynamic model were superior to the other models or estimators, even in the presence of a weak effect. In other words, if history matters then a dynamic model remains a better choice when compared to the other options. However, it is not appropriate in the presence of non-stationary data or if the model residuals are too strongly autocorrelated. Another interesting conclusion from this study was the fact that a large number of observations were not required for obtaining good coefficient estimates when using dynamic regression. Even with as few as 50 repeated measures, simple dynamic models produced good estimates and in the presence of autocorrelated residuals, these estimates were better for modest, rather than large sample sizes.

Although the use of this method is not yet common in behavioural research, dynamic regression has been previously suggested as an adequate approach to model n-of-1 data in clinical research. In their guide for the design and implementation of n-of-1 trials, Kravitz et al. (2014) briefly present dynamic models as an appropriate statistical approach to analyse autocorrelated data from n-of-1 experimental studies.

In this paper, we describe in detail the use of dynamic modelling and generalise its description to include analysis of data from observational studies. We also provide an illustrative example of the use of dynamic regression to analyse data from an n-of-1 study in health behaviour research and provide the necessary supplementary materials for the replication of the analysis.

### An illustrative example: physical activity

World Health Organisation guidelines recommend that adults should engage in at least 150 minutes of moderate (or 75 minutes of vigorous) intensity aerobic physical activity (PA) per week, performed in bouts of PA lasting at least 10 minutes in duration (World Health Organization, 2010). Identifying predictors of PA using n-of-1 studies seems a natural solution to study individual behaviour and to aid the development of individualised interventions.

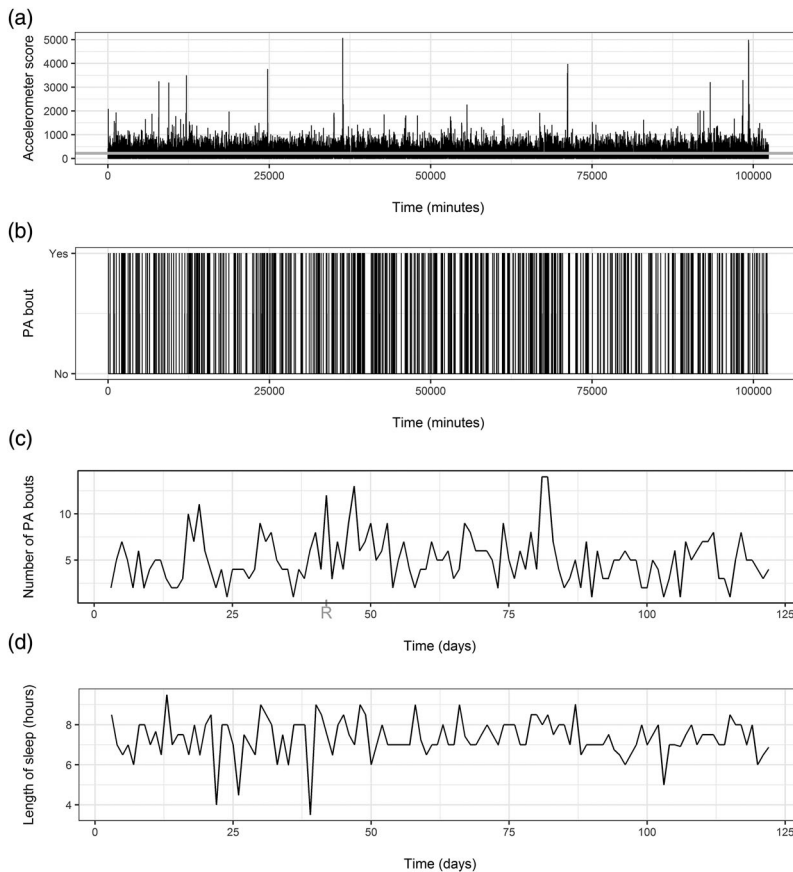
One of the latest trends in PA research is the use of fitness trackers with integrated movement detectors. In research, more sophisticated electronic devices are used to collect objectively measured accelerometry data alongside self-reported questionnaires that can be time-stamped and answered in specific contexts (e.g., Ecological Momentary Assessment). This leads to the collection of within-individual real-time data in a participant's natural environment (Shiffman, Stone, & Hufford, 2008), and can capture planned or unplanned ecological events.

McDonald, Vieira, O'Brien, White, and Sniehotta (2016) present a series of novel n-of-1 studies that intended to explore PA change during the retirement transition. The data derived from one participant from this study are represented in Figure 1. The participant being monitored wore the device for 120 days. The device sampled activity by assigning a raw score of movement performed within 60-second epochs (Figure 1(a)). Assuming that a PA 'bout' was recorded when there was a continuous string of activity corresponding to  $\geq 217$  raw counts per minute (as found in Eslinger et al., 2011 and Hickey et al., 2016), lasting at least 10 minutes, it is possible to determine the exact minute when a bout of PA was initiated (Figure 1(b)). This information can be further transformed into a daily summary of the number of PA bouts (Figure 1(c)).

In terms of predictors, the potential variables need to vary over time (e.g., the occupational state of the participant has no information if it remains the same throughout the follow-up period). In this illustrative example, the investigators were interested in exploring the relationship between sleep and PA. Therefore, the number of hours of sleep in the previous night was also collected (Figure 1(d)). One advantage of using an individualised study design is the possibility of including personalised variables applicable specifically to the individual in study (McDonald, Araújo-Soares, & Sniehotta, 2016). In this example, we will account for the fact that the participant started retirement 42 days after beginning the study, as it may potentially impact the levels of PA. The participant perceived their partner to have a significant influence on their PA behaviour. Therefore, the participant rated (on a scale of 0–1) the influence of their partner on a daily basis in response to the question 'how much did your partner influence your PA today?'

### Defining the primary outcome

Special consideration needs to be given to the distributional characteristics of the available data. In the illustrative example, there were two candidate transformations of the raw data for the PA outcome: the binary data which identifies if a continuous string of activity of at least 10 minutes was initiated at any given time, and the number of PA bouts per day. Given its binary structure, a logistic dynamic regression would be the natural choice for the analysis of the former variable, while a Poisson or even a linear dynamic regression would seem appropriate for the latter outcome, given it is count data and a large sample size is available. However, prior exploratory analysis showed that the number of PA bouts per day did not have Poisson distribution and was not normally distributed for this participant.<sup>1</sup>



**Figure 1.** (a) Example of time-series data obtained in an n-of-1 study targeting PA and length of sleep. The follow-up time was 120 days. This figure shows the raw score of movement performed within 60-seconds for each minute when the accelerometer was worn. The grey horizontal line represents the raw movement score threshold (217) used to identify PA. (b) Time at which bouts of PA were initiated, assuming a continuous string of activity is at least 10 minutes of  $\geq 217$  raw counts per minute. (c) Daily count of the number of PA bouts. Retirement day is identified in the x-axis. (d) Number of hours of sleep the previous night.

### Formally describing dynamic regression models

Dynamic regression models have been identified as an appropriate approach to express and model the behaviour of each individual over time (Kravitz et al., 2014; Schmid, 2001). Assuming that time  $t$  is a discrete variable such that  $t = 0, 1, \dots, T$ , where  $T$  is the total number of minutes the individual wore the device, then the observations of PA bouts form a subject-specific binary process  $Y_1, \dots, Y_T$ , where  $Y_t = 1$  if the participant initiates a PA bout at minute  $t$ ,  $Y_t = 0$  otherwise (Figure 1(b)). This binary variable corresponds to the dependent variable in the model.

In dynamic regression, two sets of covariates,  $X_t$  and  $D_t$ , both of which vary over time, are considered. The first,  $X_t$ , describes exogenous conditions such as trend over time, day of the week (weekend, workday), period of day (morning, afternoon and evening), together with endogenous covariates such as retirement, partner's influence and length of sleep (hours), which are specific to the individual. The model can be further adjusted for any unplanned events reported throughout the follow-up and that can potentially affect the outcome, by including covariates representing the event (e.g., an indicator variable representing the occurrence of leg injuries from a car accident).

The second,  $D_t$ , are dynamic covariates constructed to summarise the history up to  $t$  of responses  $Y_1, \dots, Y_T$ , of the individual. These may depend on an arbitrary way on the 'past'  $F_t$  which is the complete history of response and covariates to time  $t$ .

As we have a binary outcome, we will consider a logistic regression. We chose to describe the effect size as the probability of initiating a PA bout at minute  $t$ , instead of the commonly used odds ratio. Not only is it easily interpretable and meaningful, it conveniently allows deriving the same probability for different time periods (e.g., per 15 minutes or per hour). Therefore, we will describe the generic logistic model:

$$P(Y_t = 1|F_{t-1}) = \text{expit}(\alpha + \beta_0 X_t),$$

where  $P(Y_t = 1|F_{t-1})$  is the probability of a PA bout,  $t$  corresponds to the chronological ordering of observations,  $\alpha$  is the constant term and  $\text{expit}(\lambda) = \exp(\lambda)/(1 + \exp(\lambda))$  is the inverse-logit function. Dynamic modelling adjusts for autocorrelation by conditioning on the past so that the response at time  $t$  is a function of the response at time  $t - 1$  or earlier times. While in dynamic regression we can select the best function  $D_t$  of the past (see example below), conventional autoregressive models only include the previous response directly in the model. A lag 2 dynamic model would be

$$P(Y_t = 1|F_{t-1}) = \text{expit}(\alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \rho_1 Y_{t-1} + \rho_2 Y_{t-2}),$$

but the model could be extended to include as many lagged response as needed, or summaries of them such as the mean value over a previous window. In other words, we try to use the past to explain the future. Since what is in the past changes as time proceeds, it is a dynamic covariate.

### Does outcome $Y$ change over time?

As it is, this model makes the, often unrealistic, assumption that the outcome is not changing systematically with time. However, the amount of PA might exhibit both short-term and long-term patterns of change over time. PA might be increasing or decreasing linearly, or not be changing at all. It might have been highly variable in some periods or remained relatively constant until some new factor came into play. Whatever its pattern, the longer term pattern of change is usually referred to as 'trend'.

If the study duration and measurement frequency are sufficient to differentiate the trend from noise, it is possible to model the time trend by introducing time in the model. For example,

$$P(Y_t = 1|F_{t-1}) = \text{expit}(\alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \delta t),$$

where  $\delta$  represents the slope of the time trend. Exponential, quadratic or higher order polynomial terms (e.g., fractional polynomials) can be used when the time trend is not linear. However, they might consume additional degrees of freedom without significantly increasing the explanatory power.

### Estimating periodicity effects

The more repeated measurements in a time-series, the more possible it is to capture periodic patterns. For example, PA patterns are expected to differ between period of the day (morning, afternoon or evening), weather seasons or when comparing weekends and workdays. These periodicity effects can influence both the outcome and predictors of interest. In order to avoid confounding the periodicity effects with those of the predictors of interest, we need to explicitly control for the period in which the measurement is observed. This is achieved by including the relevant variables along with the other predictors. Let  $S_t$  be the set of variables describing the periodicity effects in the time-series. Then, our final dynamic model would be

$$P(Y_t = 1|F_{t-1}) = \text{expit}(\alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \theta S_t + \delta t),$$



where  $\theta$  corresponds to the periodicity effect. This way we are able to simultaneously obtain more precise estimates of both periodicity and the effects of the other predictors.

### Important assumptions when using dynamic modelling

In order to exploit n-of-1 data, three important assumptions are made. The first is that the effect of the past is captured through the exogenous, endogenous and dynamic covariates, i.e.,

$$P(Y_t = 1|F_{t-1}) = P(Y_t|X_{t-1}, D_{t-1}).$$

A misspecification of the lag structure (e.g., when the lag is one day but is defined as one week instead) might result in a sign reversal of the contemporaneous effect (Vaisey & Miles, 2017). The second important assumption is stationarity, i.e., the dynamics do not change over time, so that the coefficients ( $\beta$ ,  $\rho$ ,  $\theta$ ,  $\delta$ ) are time constant. A third assumption is that the dynamic covariates do not lie on the causal pathways between the exogenous/endogenous covariates  $X_t$  and the response  $Y_t$ .

### Extending the model

In this example, no intervention was introduced. However, when considering individualised n-of-1 experimental studies such as randomised or counterbalanced designs (e.g., ABBABAAB), careful consideration is needed, not only to account for period and treatment effects, but also carryover effects. See Kravitz et al. (2014) for more details on the statistical and analytical considerations of n-of-1 experimental studies.

Dynamic regression can accommodate different types of outcomes by using an appropriate link function when building the model. Logistic regression is appropriate for a binary outcome like in our example and it can be extended to an unordered or ordered categorical logistic regression if the outcome has more than two unordered or ordered categories, respectively. A linear regression would be appropriate for continuous data with normally distributed errors, while a Poisson regression is usually adequate to model count data.

### Is variable $Y$ a function of variable $X$ ?

Returning to the illustrative example, Table 1 lists the variables included in the model. The model assumes a binary outcome representing the initiation (or not) of a bout of PA in each minute recorded by the activity tracker. Having controlled for the acknowledged sources of confounding, focus is now on the main research question: does the number of hours the individual sleeps influence how physically active the participant is?

Table 2 summarises the effect sizes (probability of initiating a PA bout per minute) and statistical significance of all variables included in the model. Assuming a cut-off  $p$ -value of .05, time has a statistically significant linearly decreasing effect on PA. Figure 2 shows that the probability of engaging in a bout of PA per hour also decreased during the weekend, although not significantly, and in the evening (when compared with morning and afternoon). Autocorrelation in the outcome is significant when accounting for the number of PA bouts in the previous two hours, but not for the total number of bouts one or two days before.

While retirement and partner's influence have a non-statistically significant effect on PA, both the number of hours the participant slept the previous night and the hours of sleep two nights before are statistically significant. Figure 3 graphically represents the probability of a bout of PA per hour, given how many hours the participant slept the night before and two nights before, considering that the participant slept on average 7.5 hours per night (range: 3.5–9.5). We observe that if the participant sleeps the average number of hours two nights before, the probability of engaging in PA will range

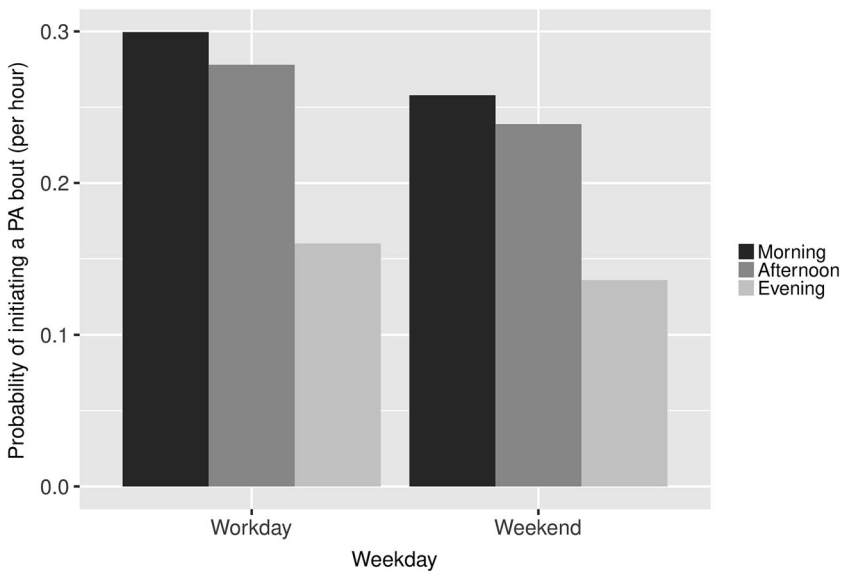
**Table 1.** Variables included in the dynamic model.

Type of variable		Description
Outcome		Initiation of a PA bout at each minute (yes/no)
Exogenous	Time	Number of days since starting the study (we assume a linear trend)
	Periodicity	Weekend/work day
Endogenous		Period of the day (afternoon/morning/evening)
		Number of hours of sleep – previous night (lag 0)
		Retirement
Dynamic	Predictor-related	Influence of partner in PA
		Number of hours of sleep – previous night (lag 1)
	Outcome-related	Number of hours of sleep – two nights before (lag 2)
		Number of PA bouts – previous two hours
		Number of PA bouts – previous day (lag 1)
	Number of PA bouts – two days before (lag 2)	

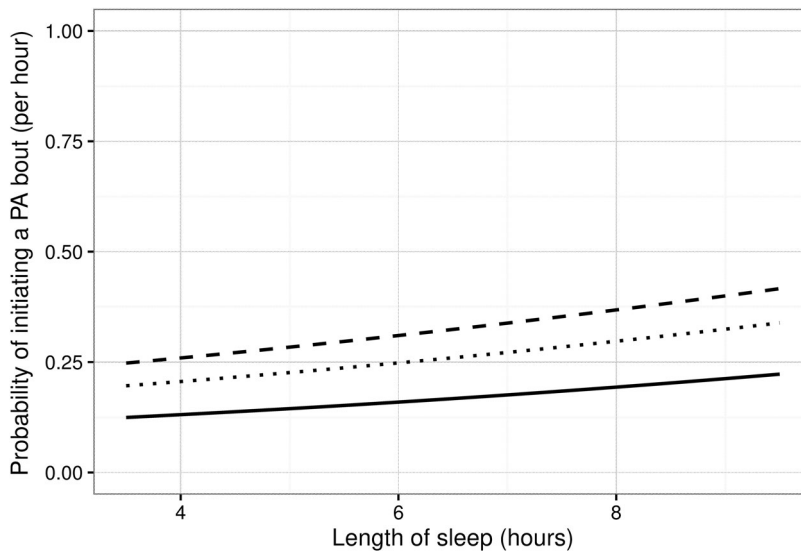
**Table 2.** Multivariable associations between predictors and PA.

Predictor	$\beta$ (SE) <sup>a</sup>	p-Value
Time trend – linear	–0.004 (0.002)	.05
Weekday (ref: work day)		
Weekend	–0.18 (0.11)	.09
Period of day (ref: afternoon)		
Morning	0.09 (0.09)	.33
Evening	–0.63 (0.12)	<.001
Number of PA bouts previous two hours	0.10 (0.04)	.01
Number of PA bouts – previous day	0.01 (0.02)	.58
Number of PA bouts – two days before	0.002 (0.02)	.92
Retirement	0.25 (0.15)	.11
Influence of partner	–0.70 (0.44)	.11
Sleep length	0.11 (0.05)	.03
Sleep length – previous day	0.03 (0.04)	.49
Sleep length – two days before	0.13 (0.05)	.01

<sup>a</sup>Regression coefficient and corresponding standard error (SE).



**Figure 2.** Probability of initiating a bout of PA per hour, by weekday and period of day.



**Figure 3.** Probability of initiating a PA bout per hour, given how many hours the individual slept the previous night and two nights before. Each line represents the probability of a PA bout per hour when the individual sleeps 9.5 (dashed), 7.5 (dotted) and 3.5 (solid) hours, two nights before.

from 20% to 34%, depending on how many hours of sleep reported for the previous night (dotted line). If the number of hours slept two days ago were the minimum 3.5 hours, the probability ranges between 12% and 22% (solid line). If the participant has slept the maximum number of hours two days ago, then the probability would range between 25% and more than 40% (dashed line). If the lag 2 effect is unaccounted for, we would detect an increase of 14% in the probability of being physically active, when comparing sleeping the minimum (3.5) and maximum (9.5) number of hours the previous night (dotted line). When acknowledging how much the participant slept two nights before, the probability of engaging in PA on the current day can range from 12% to 40%, almost a 30% increase. This example shows the importance of modelling the dynamic effect of variables, (i.e., the effect of the past on the future). It also shows that methods which intend to remove autocorrelation prior to the analysis (e.g., pre-whitening, as used by Hobbs, Dixon, Johnston, & Howie, 2013; Kwasnicka et al., 2017; Quinn et al., 2013) need to be used carefully as they might remove an existing effect.

### What about sample size?

The analysis of the PA data presented in this paper is for descriptive and modelling purposes rather than to test a specific null hypothesis. However, calculating a sample size for n-of-1 studies is challenging as concerns exist on whether there is accurate prior information on likely variances and dependence structures for sample size calculation. More importantly, it is often difficult to define a null or alternative hypothesis or state the clinically important differences needed for power calculations. Even if the latter is achieved, the final sample size will also depend on practical considerations related to feasibility and type of measurement, for example, an individual's ability and willingness to record data more than once a day or for a long period of time. Clearly, this is an area in need of further discussion.

### In conclusion

N-of-1 studies offer an opportunity to explore within-individual variability and develop individualised interventions in health psychology, but the apparent complexity and uncertainty around its statistical

analysis may have limited the use of these study designs in the past. In this article we have described the statistical and analytical considerations of n-of-1 studies, emphasising the time-series structure of the data. Several methodologies are available that appropriately model time-series data and provide adequate effect sizes. However, their complexity may have limited their usability. Dynamic modelling is a robust and commonly used method to analyse time-series data. It adjusts for autocorrelation and models the dynamic effects by incorporating the dependence of future on past. This is achieved by including lagged covariates, representing the past history of the predictors of interest and the outcome, in conventional multivariable regression models, therefore avoiding the use of more complex estimation methods as, for example, when using autoregressive models. However, as dynamic modelling assumes that the effect of the past is captured through the dynamic covariates, these lagged variables need to be carefully chosen to represent the past appropriately. It further assumes that the dynamics do not change over time, which means it is applicable only for stationary time-series data. Nevertheless, dynamic modelling is an intuitive, robust and flexible statistical approach with potential to strengthen the quality of evidence generated in future individualised studies.

## Note

1. Further transformations could be used in an attempt to normalise the data but it would further complicate the presentation of the example. We acknowledge that a continuous and normally distributed outcome would facilitate the description of the model but we consider it to be of value to exemplify an analysis of a real-world n-of-1 study, where outcomes are more often than not, non-normally distributed.

## Acknowledgements

The authors thank Dr Kirby Sainsbury for the insight given on the challenges related with analysing n-of-1 data and statistician Dr Kate Best for her comments and help with drafting some of the ideas expressed in the manuscript. The authors thank Professor Martin White and Dr Nicola O'Brien for their contributions towards the conception and design of the n-of-1 study which formed the basis of the illustrative example provided in the paper. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NIHR, NHS or the Department of Health.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was part-funded by the National Institute for Health Research (NIHR) Programme Grants for Applied Research (RP-PG-1212-20009). FFS is funded by Fuse, the Centre for Translational Research in Public Health, a UKCRC Public Health Research Centre of Excellence based on funding from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration.

## Supplementary materials

An R script with syntax to generate the results and plots presented in this paper is available as supplementary materials to encourage wider use. The script also includes example syntax on how to analyse data with a normally distributed outcome as well as a Poisson distributed count outcome. The data used in the illustrative example is available via <https://doi.org/10.5281/zenodo.580028>.

## References

- Araújo, A., Julious, S., & Senn, S. (2016). Understanding variation in sets of n-of-1 trials. *PLoS One*, *11*(12), e0167167.
- Bagian, J. P., King, B. J., Mills, P. D., & McKnight, S. D. (2011). Improving RCA performance: The cornerstone award and the power of positive reinforcement. *BMJ Quality & Safety*, *20*(11), 974–982.
- Barlow, D., Nock, M., & Hersen, M. (2008). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Upper Saddle River: Pearson Education.
- Barr, C., Marois, M., Sim, I., Schmid, C. H., Wilsey, B., Ward, D., ... Kravitz, R. (2015). The PREEMPT study – evaluating smart-phone-assisted n-of-1 trials in patients with chronic pain: Study protocol for a randomized controlled trial. *Trials*, *16*, 200.
- Bombom, O., & van der Laan, M. J. (2008). Analyzing sequentially randomized trials based on causal effect models for realistic individualized treatment rules. *Statistics in Medicine*, *27*, 3689–3716.
- Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O'Neil, P. (2008). Clinical practice as natural laboratory for psychotherapy research: A guide to case-based time-series analysis. *American Psychologist*, *63*(2), 77–95.
- Box, G. E. P., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.
- Cheung, Y. K., Chakraborty, B., & Davidson, K. W. (2015). Sequential multiple assignment randomized trial (SMART) with adaptive randomization for quality improvement in depression treatment program. *Biometrics*, *71*(2), 450–459.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed., pp. 219–223). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: The new medical research council guidance. *BMJ*, *337*, a1655.
- Dempsey, W., Liao, P., Klasnja, P., Nahum-Shani, I., & Murphy, S. A. (2015). Randomised trials for the Fitbit generation. *Significance (Oxford)*, *12*(6), 20–23.
- Duan, N., Kravitz, R. L., & Schmid, C. H. (2013). Single-patient (n-of-1) trials: A pragmatic clinical decision methodology for patient-centered comparative effectiveness research. *Journal of Clinical Epidemiology*, *66*(8), S21–S28.
- Esliger, D. W., Rowlands, A. V., Hurst, T. L., Catt, M., Murray, P., & Eston, R. G. (2011). Validation of the GENEA accelerometer. *Medicine & Science in Sports & Exercise*, *43*(6), 1085–1093.
- Gabler, N. B., Duan, N., Vohra, S., & Kravitz, R. L. (2011). N-of-1 trials in the medical literature: A systematic review. *Medical Care*, *49*(8), 761–768.
- Glass, G. V., Willson, V. L., & Gottman, L. M. (1975). *Design and analysis of time-series experiments*. Boulder: Colorado Associated University Press.
- Henderson, R., Ansell, P., & Alshibani, D. (2010). Regret-regression for optimal dynamic treatment regimes. *Biometrics*, *66*, 1192–1201.
- Hendry, D. F. (1995). *Dynamic econometrics*. Oxford: Oxford University Press.
- Hickey, A., Newham, J., Slawinska, M. M., Kwasnicka, D., McDonald, S., Del Din, S., ... Godfrey, A. (2016). Estimating cut points: A simple method for new wearables. *Maturitas*, *83*, 78–82.
- Hinich, M. J. (1982). Testing for Gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*, *3*, 169–176.
- Hobbs, N., Dixon, D., Johnston, M., & Howie, K. (2013). Can the theory of planned behaviour predict the physical activity behaviour of individuals? *Psychology & Health*, *28*, 234–249.
- Johnston, D. W., & Johnston, M. (2013). Useful theories should apply to individuals. *British Journal of Health Psychology*, *18*, 469–473.
- Kazdin, A. E. (2011). *Single-case research designs* (2nd ed.). New York, NY: Oxford University Press.
- Keele, L., & Kelly, N. J. (2006). Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political Analysis*, *14*, 186–205.
- Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., & Murphy, S. A. (2015). Micro-randomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology*, *34*, 1220–1228.
- Kravitz, R. L., Duan, N. (Eds.), & the DEClDE Methods Center N-of-1 Guidance Panel (Duan, N., Eslick, I., Gabler, N. B., Kaplan, H. C., Kravitz, R. L., Larson, E. B., ... Vohra, S.). (2014). *Design and implementation of n-of-1 trials: A user's guide* (AHRQ Publication No. 13(14)-EHC122-EF). Rockville, MD: Agency for Healthcare Research and Quality.
- Kravitz, R. L., Paterniti, D. A., Hay, M. C., Subramanian, S., Dean, D. E., Weisner, T., ... Duan, N. (2009). Marketing therapeutic precision: Potential facilitators and barriers to adoption of n-of-1 trials. *Contemporary Clinical Trials*, *30*(5), 436–445.
- Kwasnicka, D., Dombrowski, S. U., White, M., & Sniehotta, F. F. (2017). N-of-1 study of weight loss maintenance assessing predictors of physical activity, adherence to weight loss plan and weight change. *Psychology & Health*, *32*, 686–708. doi:10.1080/08870446.2017.1293057
- Law, L. M., Edirisinghe, N., & Wason, J. M. (2016). Use of an embedded, micro-randomised trial to investigate non-compliance in telehealth interventions. *Clinical Trials*, *13*(4), 417–424.
- Liao, P., Klasnja, P., Tewari, A., & Murphy, S. A. (2016). Sample size calculations for micro-randomized trials in mHealth. *Statistics in Medicine*, *35*(12), 1944–1971.

- Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., & Schork, N. J. (2011). The n-of-1 clinical trial: The ultimate strategy for individualizing medicine? *Personalized Medicine*, 8(2), 161–173.
- Mao, X., & Cheung, Y. K. (2016). Sequential designs for individualized dosing in phase I cancer clinical trials. *Contemporary Clinical Trials*. pii: S1551-7144(16)30256-7.
- McDonald, S., Araújo-Soares, V., & Sniehotta, F. F. (2016). N-of-1 randomised controlled trials in health psychology and behavioural medicine: A commentary on Nyman et al., 2016. *Psychology & Health*, 31(3), 331–333.
- McDonald, S., Quinn, F., Vieira, R., O'Brien, N., White, M., Johnston, D. W., & Sniehotta, F. F. (2017). The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: A systematic literature overview. *Health Psychology Review*. doi:10.1080/17437199.2017.1316672
- McDonald, S., Vieira, R., O'Brien, N., White, M., & Sniehotta, F. F. (2016). Does physical activity and sedentary behavior change during the retirement transition? Findings from a series of novel n-of-1 natural experiments. *International Journal of Behavioral Medicine*, 23, S261–S261.
- McKnight, S. D., McKean, J. W., & Huitema, B. E. (2000). A double bootstrap method to analyze linear models with autoregressive error terms. *Psychological Methods*, 5(1), 87–101.
- Mills, T. C. (1990). *Time series techniques for economists*. Cambridge: Cambridge University Press.
- Nash, M. R., Borckardt, J. J., Abbasa, A., & Gray, E. (2011). How to conduct and statistically analyze case-based time series studies, one patient at a time. *Journal of Experimental Psychopathology*, 2(2), 139–169.
- Naughton, F., & Johnston, D. (2014). A starter kit for undertaking n-of-1 trials. *The European Health Psychologist*, 16(5), 196–205.
- O'Brien, N., Philpott-Morgan, S., & Dixon, D. (2016). Using impairment and cognitions to predict walking in osteoarthritis: A series of n-of-1 studies with an individually tailored, data-driven intervention. *British Journal of Health Psychology*, 21, 52–70. doi:10.1111/bjhp.12153
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, 35(4), 303–322.
- Perdices, M., & Tate, R. L. (2009). Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognized and undervalued? *Neuropsychological Rehabilitation*, 19(6), 904–927.
- Person Centred Care/Coalition for Collaborative Care/Medical Directorate. (2016). *NHS England personalised care & support planning handbook – core information*. Retrieved from [www.mheducation.co.uk/openup/chapters/9780335246267.pdf](http://www.mheducation.co.uk/openup/chapters/9780335246267.pdf)
- Pevehouse, J. C., & Brozek, J. D. (2008). *Time-series analysis. The Oxford handbook of political methodology*. Oxford: Oxford University Press.
- Prais, S. J., & Winsten, C. B. (1954). *Trend estimators and serial correlation* (Cowles Commission Discussion Paper No. 383). Chicago, IL.
- Quinn, F., Johnston, M., & Johnston, D. W. (2013). Testing an integrated behavioural and biomedical model of disability in n-of-1 studies with chronic pain. *Psychology & Health*, 28(12), 1391–1406.
- Rothwell, P. M. (2005). External validity of randomised controlled trials: 'To whom do the results of this trial apply?' *The Lancet*, 365(9453), 82–93.
- Rusticelli, E., Ashley, R. A., Dagum, E. B., & Patterson, D. M. (2008). A new bispectral test for nonlinear serial dependence. *Econometric Reviews*, 28(1–3), 279–293.
- Sainsbury, K., Walburn, J., Vieira, R., Sniehotta, F. F., Weinman, J., Sarkany, R., & Araujo-Soares, V. (2016). Using n-of-1 methodology to inform the development of individualized, evidence-based interventions for patients with xeroderma pigmentosum. *International Journal of Behavioral Medicine*, 23, S46–S46.
- Schmid, C. H. (2001). Marginal and dynamic regression models for longitudinal data. *Statistics in Medicine*, 20(21), 3295–3311.
- Shadish, W. (2014). Statistical analyses of single-case designs: The shape of things to come. *Current Directions in Psychological Science*, 23(2), 139–146.
- Shamseer, L., Sampson, M., Bukutu, C., Schmid, C., Nikles, J., Tate, R., ... Vohra, S. (2015). CONSORT extension for reporting n-of-1 trials (CENT) 2015: Explanation and elaboration. *BMJ*, 350, h1793.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32.
- Sniehotta, F. F., Pesseau, J., Hobbs, N., & Araújo-Soares, V. (2012). Testing self-regulation interventions to increase walking using factorial randomized n-of-1 trials. *Health Psychology*, 31(6), 733–737.
- Swaminathan, H., Rogers, H. J., & Horner, R. H. (2014). An effect size measure and Bayesian analysis of single-case designs. *Journal of School Psychology*, 52(2), 213–230.
- Vaisey, S., & Miles, A. (2017). What you can – and can't – do with three-wave panel data. *Sociological Methods and Research*, 46(1), 44–67.
- Velicer, W. F., & Fava, J. L. (2003). Time Series Analysis. In J. Schinka & W. F. Velicer (Eds.), *Research Methods in Psychology. Handbook of Psychology* (Vol. 2, pp. 581–606). New York: John Wiley & Sons.
- World Health Organization. (2010). *Global recommendations on physical activity for health*. Retrieved from [http://www.who.int/dietphysicalactivity/factsheet\\_recommendations/en/](http://www.who.int/dietphysicalactivity/factsheet_recommendations/en/)