INTEGRATING SLEEP, SEDENTARY BEHAVIOUR, AND PHYSICAL ACTIVITY RESEARCH IN THE EMERGING FIELD OF TIME-USE EPIDEMIOLOGY: DEFINITIONS, CONCEPTS, STATISTICAL METHODS, THEORETICAL FRAMEWORK, AND FUTURE DIRECTIONS

Željko Pedišić, Dorothea Dumuid, and Timothy S. Olds

1Institute of Sport, Exercise and Active Living (ISEAL), Victoria University, Melbourne, Australia
2School of Health Sciences, University of South Australia, Adelaide, Australia

Abstract:

Nearly 70 years of sleep, sedentary behaviour, physical activity, and time-use research has led to the recent development of time-use epidemiology. To conceptualise the emerging research field and provide a framework for its further development, this paper defines its position among the established branches of science, explains its main concepts and defines associated terms, recommends suitable data analysis methods, proposes a theoretical model for future research, and identifies key research questions. Time-use epidemiology is defined as the study of determinants, incidence, distributions, and effects of health-related time-use patterns in populations and of methods for preventing unhealthy time-use patterns and achieving the optimal distribution of time for population health. As a theoretical model for future studies, this paper proposes the Framework for Viable Integrative Research in Time-Use Epidemiology (VIRTUE framework), acknowledging the compositional nature of time-use data and incorporating research on: 1) methods in time-use epidemiology; 2) outcomes of health-related components of time use; 3) optimal time-use balance and its prevalence in populations; 4) determinants and correlates of health-related components of time use; and 5) effectiveness of time-use interventions. It is likely that in total more deaths worldwide can be attributed to unhealthy time use than to smoking or obesity, potentially making it the most relevant modifiable behavioural and lifestyle risk factor of our time. We hope that governments and leading health organisations will recognise enormous importance of healthy time use, and provide adequate support for future research in time-use epidemiology.

Key words: compositional data analysis, VIRTUE framework, optimal time-use balance, time-use interventions, Activity Balance Model

Introduction: towards time-use epidemiology

Every given period of time people spend sleeping, in sedentary behaviour, in quiet standing, or in physical activity. A plethora of evidence strongly indicates that these components of time use are related to a number of health outcomes. For example, inadequate duration of sleep, high sitting time, and insufficient time spent in physical activity all seem to increase the risk of all-cause mortality, cardiovascular disease, metabolic syndrome, type 2 diabetes, and some types of cancer (Cappuccio, Cooper, Delia, Strazzullo, & Miller, 2011; Cappuccio, D’Elia, Strazzullo, & Miller, 2010a, 2010b; de Rezende, Lopes, Rey-López, Matsudo, & Luiz, 2014; Lee, et al., 2012; Xi, He, Zhang, Xue, & Zhou, 2014; Zhao, et al., 2013). Epidemiologists have traditionally investigated these behaviours in isolation from each other, and have deemed them independent health risk factors.

Insufficient physical activity has been identified as a potential health risk factor more than 60 years ago (Morris, Heady, Raffle, Roberts, & Parks, 1953). Most studies in this area have been focused on moderate-to-vigorous physical activity (MVPA). The first epidemiological studies on sleep can be dated to late seventies of the last century (Bixler, Kales, Soldatos, Kales, & Healey, 1979). A rapid development of sedentary behaviour research started nearly two decades ago after researchers had called for making a clear distinction between physical inactivity, i.e. lack of MVPA, and sedentary
behave, i.e. sitting or reclining with low energy expenditure (Marshall, Biddle, Sallis, McKenzie, & Conway, 2002; Owen, Bauman, & Brown, 2009; Owen, Leslie, Salmon, & Fotheringham, 2000; Pate, O’Neill, & Lobelo, 2008). More recently, studies have also indicated that quiet standing and light physical activity (LPA) may be associated with health outcomes (Halim & Omar, 2011; Healy, et al., 2007; Katzmarzyk, 2014; van der Ploeg, et al., 2014; Waters & Dick, 2015).

In 2007, Tremblay and colleagues suggested that an appropriate balance between sleep, physical inactivity, and physical activity is needed for optimal health (Tremblay, Esliger, Tremblay, & Colley, 2007), and that, therefore, public health guidelines should integrate recommendations on all these behaviours. Mekary and colleagues introduced the isocostal substitution procedure to the field of physical activity and sedentary behaviour, by demonstrating how re-allocation of time between these two behaviours is associated with weight change (Mekary, Willett, Hu, & Ding, 2009). The same procedure was later used to examine associations of time re-allocations between sleep, sedentary behaviour, and physical activity with self-regulatory behaviour and executive functioning (Fanning, et al., 2017), symptoms of fatigue (Vallance, Buman, Lynch, & Boyle, 2017), quality of life (Vallance, et al., 2017), body mass index (BMI) (Boyle, Vallance, Buman, & Lynch, 2017; Huang, Wong, He, & Salmon, 2016), waist circumference (Boyle, et al., 2017), cardiovascular disease risk biomarkers (Buman, et al., 2014), and all-cause mortality risk (Stamatakis, et al., 2015).

In 2014, Pedišić called attention to statistical inadequacy of the isocostal substitution model and other traditionally used multivariate methods when analysing associations of sleep, sedentary behaviour, quiet standing, LPA and MVPA with health outcomes, and proposed a change of the research paradigm towards an integrated approach. In the theoretical framework called the Activity Balance Model, Pedišić (2014) suggested that, instead of being treated as independent risk factors, sleep, sedentary behaviour, quiet standing, and physical activity should be analysed as parts of a time-use composition using compositional data analysis. The potential importance of applying this new concept to advance physical activity and sedentary behaviour research was soon recognised in the field (Biddle, García Bengoechea, & Wiesner, 2017; Chaput, Carson, Gray, & Tremblay, 2014; Chaput, Saunders, & Carson, 2017; Kang & Rowe, 2015; Lynch & Owen, 2015; Wijndaele & Healy, 2016), and a range of empirical studies were conducted using the suggested analytical approach (Carson, Tremblay, Chaput, & Chastin, 2016; Chastin, Palarea-Albaladejo, Donjtje, & Skelton, 2015; Dumuid, et al., 2016; Dumuid, Olds, Lewis, et al., 2017; Dumuid, Olds, Martin-Fernández, et al., 2017; Dumuid, Stanford, et al., 2017; Fairclough, et al., 2017).

In 2016, Tremblay and colleagues led the development of the Canadian national public health guidelines for children and adolescents, which for the first time included recommendations encompassing sleep, sedentary behaviour, LPA, and MVPA (Tremblay, et al., 2016). The development of the guidelines was informed by a range of studies acknowledging the integrative approach to all components of the 24-hour time-use continuum. In 2017, the New Zealand Ministry of Health followed the Canadian example and released joint recommendations for sleep, sedentary behaviour, and physical activity for children and young people (New Zealand Ministry of Health, 2017). Public health authorities in Australia and Denmark are now considering issuing similar national guidelines on the optimal time use. The integrative approach to sleep, sedentary behaviour and physical activity research has attracted great attention among scholars and is gaining momentum in the research community. This is evidenced by dedicated sessions, workshops and/or individual presentations at major conferences on physical activity and health, such as the International Conference on Diet and Activity Methods (ICDAM), International Society for Physical Activity and Health (ISPAH) Biennial Congress, the Annual Meeting of the International Society for Behavioral Nutrition and Physical Activity (ISBNPA), and the HEPA Europe Conference. Increased availability of 24-hour continuous data on sleep, sedentary behaviour, and physical activity from accelerometer-based studies and time-use surveys adds to the potential for a fast and widespread adoption of the new research paradigm in the years to come (Pedišić & Bauman, 2015; van der Ploeg, et al., 2010).

All these recent developments (Figure 1) have led to the establishment of a new field in public health research that can be best labelled as time-use epidemiology. To conceptualise the emerging research field of time-use epidemiology and provide a framework for its further development, we here with define its position among the established branches of science, explain its main concepts and define associated terms, recommend suitable data analysis methods, propose a theoretical model for future research, and identify key research questions.

Position of time-use epidemiology among the established scientific fields

Time-use research

Time-use research is a multidisciplinary field of study focused on investigating individual and societal allocation of time during a given period, usually a day or a week, and factors influencing time-use choices. Time-use studies have commonly
explored distribution of time across different categories, such as contracted time, committed time, necessary time, and free time (Aas, 1982). Typical research questions in this area have been: “How do people balance their work time and leisure time?”, “How do people schedule their work?”, “Which activities do people choose as part of their leisure time?”, “What is the amount of time spent doing housework?”, “How do gender and other sociodemographic factors affect individual’s time use?”, “What is the influence of urban design on individuals’ time-use choices?”, “How do time-use patterns affect individuals’ social connectedness, within and outside their households?”, “What are the economic implications of time distribution?”, “How do time-use choices affect social and overall well-being?”. Important areas of time-use research have also included development and validation of survey methods, particularly time-use diaries and questionnaires, and time-use classification systems (Merz, 2002).

**Epidemiology and behavioural epidemiology(ies)**

Epidemiology is a multidisciplinary field of study focused on investigating incidence and distribution of health-related events, states, and processes in populations, their determinants, correlates, and effects, and ways to prevent and control health problems at the population level (International Epidemiological Association, 2014; Lilienfeld, 1978). Early epidemiological research was primarily focused on infectious diseases. However, the growing burden of non-communicable diseases (NCDs) over the last century has driven the scope of epidemiology to also include NCDs and their behavioural risk factors such as physical inactivity, smoking, alcohol intake, and unhealthy diet. During the 1970s, the term “behavioural epidemiology” was introduced to describe the evolving field of health behaviour research. In 2000, Sallis and colleagues presented a detailed conceptualisation of behavioural epidemiology as a sub-discipline of epidemiology (Sallis, Owen, & Fotheringham, 2000). Many epidemiologists have specifically investigated one particular health behaviour, leading to the formation of separate epidemiological subfields, such as nutritional epidemiology (Michels, 2003), physical activity epidemiology (Caspersen, 1989), and sleep epidemiology (Ferrie, Kumari, Salo, Singh-Manoux, & Kivimäki, 2011).

**Time-use epidemiology**

Recent methodological advances call for an urgent integration of research on time spent sleeping, in sedentary behaviour, and in physical activity. Methodological advances include the development and validation of new measurement tools and techniques, as well as the use of compositional data analysis (CoDA) to analyse time-use data. The use of CoDA can help to identify patterns and trends in time-use behaviour, and to understand the relationships between time-use and health outcomes.
activity into a unified field of time-use epidemiology (Chastin, Palarea-Albaladejo, et al., 2015; Dumuid, Stanford, et al., 2017; Pedišić, 2014). The amounts of time spent sleeping, in sedentary behaviour, and in physical activity are components of time use, and, as such, they conceptually belong under time-use research. The establishment of time-use epidemiology has been motivated by the methodological limitations of previous studies investigating the health effects of sleep, sedentary behaviour, and physical activity (Pedišić, 2014). However, time-use epidemiology is not restricted only to these particular time-use variables. It also incorporates research on any other time-use variables that may be related to health; for example, family time, peer time, housework, leisure, child care, productivity time, and self-care (Barnes, Hoffman, Welte, Farrell, & Dintcheff, 2007; Bird & Fremont, 1991; Desha & Ziviani, 2007; Jowsey, et al., 2013; Pentland, Harvey, & Walker, 1998). Time-use epidemiology can, therefore, be defined as the study of determinants, incidence, distributions, and effects of health-related time-use patterns in populations and of methods for preventing unhealthy time-use patterns and achieving the optimal distribution of time for population health. Time-use epidemiology incorporates a great deal of but not necessarily all of sleep, sedentary behaviour, and physical activity research (Figure 2). There may be topics in these research areas that do not necessarily require integration into time-use epidemiology, for example, studies on sleep disturbances, physical activity policy research, studies on attitudes towards physical activity, or qualitative evaluations of sit-stand desk interventions. It is also important to clearly distinguish between health- and non-health-related aspects of time use. The latter are not within the scope of time-use epidemiology.

Key terms and concepts in time-use epidemiology

Health-related components of time use

Among the different time-use components that have been investigated in relation to health, by far most commonly studied have been sleep, sedentary behaviour, and physical activity. Criteria used to classify time into sleep, sedentary behaviour, and physical activity are: 1) wakefulness (awake / not awake); 2) posture (lying / sitting / standing); and 3) relative energy expenditure (≤1.5 metabolic equivalents [METs] / > 1.5 METs). Sleep is a spontaneous and reversible state of rest characterised by inhibition of voluntary muscles and sensory activity, and by reduced consciousness, responsiveness to stimuli, and interactions with the environment (Carskadon & Dement, 2011). With a relative energy expenditure of 0.95 METs, sleep is an activity at the lowest end of the intensity spectrum (Ainsworth, et al., 2011). Sedentary behaviour is characterised by wakefulness, a low energy expenditure of ≤1.5 METs, and a sitting or reclining posture (Sedentary Behaviour Research Network, 2012). Physical activity is bodily movement produced by skeletal muscles that results in energy expenditure (Caspersen, Powell, & Christenson, 1985). Physical activity is usually categorised by its intensity level, that is, the level of relative energy expenditure, into: [i] LPA (generally 1.5-3 METs); [ii] moderate-intensity physical activity (generally 3-6 METs); and [iii] vigorous-intensity physical activity (>6 METs). For research purposes, moderate-intensity and vigorous-intensity physical activity levels are often collapsed into a single category, i.e. MVPA, whilst the LPA category usually includes quiet standing. In the following text, the term LPA refers to the combined category of LPA and quiet standing. Other time-use variables that are potentially related to health may be derived by applying and combining different sets of criteria for time-use classification, for example, “social / non-social”; “voluntary / obligatory”; “at work / in transport / as part of housework / as part of self-care / as part of leisure-time”; “outdoors / indoors”, etc. There are many systems for classifying types of daily activities, and their selection reflects different research emphases. Hierarchical aggregation systems are often used for classification, whilst in some cases one system of criteria can be mapped onto another (e.g. when ascribing energy expenditure bands to classic time-use surveys). To avoid potentially lengthy phrases when denoting a time-use composition, we suggest using the code consisting of the Wingdings threeoclock symbol followed by abbreviations for its parts linked with the en dash symbol. For example, the time-use composition consisting of sleep, sedentary behaviour, LPA, and MVPA would be coded as SLP-SB-LPA-MVPA.
Compositional nature of time-use data

The amounts of time spent sleeping and in sedentary behaviour, LPA, and MVPA are mutually exclusive components of a 24-hour day or any other fixed time period. This means that, at the same time, one can engage in only one of these activities. Sleep, sedentary behaviour, LPA, and MVPA are also exhaustive components of time use, that is, their total sum always constitutes the whole 24-hour day or any other unit of time. These data are meaningfully interpretable as proportions of a whole. Proportions of time spent in these behaviours may significantly vary within and between individuals, and across different population groups and countries, but, despite that, they always add up to 1, i.e. 100%. The constant sum constraint makes the proportions of time spent sleeping, in sedentary behaviour, in LPA, and in MVPA perfectly collinear. This means that every change in time spent in one of the behaviours necessarily results in a corresponding opposite change of the time spent in one or more remaining behaviours. The sample space of such constrained data, classified in statistical literature as a case of compositional data, is vastly different from real space associated with unconstrained vectors (Aitchison, 1982, 1986; Barceló-Vidal & Martín-Fernández, 2016). Therefore, the specific mathematical properties of compositional vectors must be taken into account when dealing with time-use data.

As noted before, research on time spent sleeping, and in sedentary behaviour, LPA, and MVPA is a key pillar of time-use epidemiology, but other components of time use that are potentially associated with health should not be disregarded. Irrespective of the criteria used to classify time use, the resulting time-use composition will have the same mathematical properties as \( \mathbb{S}_D \)-LPAs-MVPA. For more detailed information about different cases of compositional data and their mathematical properties, see Barceló-Vidal and Martín-Fernández (2016).

Statistical approaches in time-use epidemiology: discerning between right and wrong

A review of 54 longitudinal studies on health outcomes of sedentary behaviour found that not even one of them adequately accounted for sleep, LPA, and MVPA (Pedišić, 2014). In our extended and updated review of literature, we found no longitudinal studies on health outcomes of sleep, sedentary behaviour, LPA, and MVPA that used an adequate analytical approach to account for compositional nature of time-use data. Therefore, results on longitudinal relationships of sleep, sedentary behaviour, LPA, and MVPA with health outcomes reported in previous studies and afterwards pooled in a number of meta-analytical reviews are potentially incorrect.

Moreover, previous conclusions about time spent sleeping, and in sedentary behaviour, LPA, and MVPA as independent health risk factors are methodologically flawed, because they ignore the true, compositional nature of time-use data. The distinctive features of compositional data make it inappropriate to analyse amounts of time spent sleeping, in sedentary behaviour, and in physical activity in isolation from each other. Because of their compositional properties and perfect collinearity, these time-use variables cannot even theoretically be considered as independent health risk factors. There have been attempts to justify analysing the relationship between a time-use variable and a health outcome without accounting for all the remaining parts of its respective time-use composition. For example, Page and colleagues suggested that in the case where physical activity is neither a common cause of both sedentary behaviour and the health outcome, nor an effect of sedentary behaviour on the causal pathway to the health outcome, then no adjustments for physical activity are needed (Page, Peeters, & Merom, 2015). However, such justifications may be misleading, because they rely on a wrong assumption that raw amounts of time spent sleeping, and in sedentary behaviour, LPA, and MVPA are vectors in real space. Moreover, even if time-use data were not compositional (although they undoubtedly always are), statistical simulations suggest that theoretical confounders should be adjusted for regardless of the empirical evidence of their confounding effects in the given dataset (Lee, 2014).

Recognising the right sample space of the data is essential for selecting the appropriate statistical model. This is because group operations play an important role in formulas for statistical data analysis, and operations that are valid for one sample space are not necessarily applicable to other sample spaces. It is widely accepted in the mathematical community that compositions are equivalence classes and their true sample space is a quotient space (Barceló-Vidal & Martín-Fernández, 2016). A representative of the sample space of compositional vectors is D-part simplex, where \( D \) is the number of parts of the composition. Basic operations in D-part simplex are perturbation and powering, and they fundamentally differ from their analogous operations in real space, namely addition and scalar multiplication. Accordingly, distances and angles between vectors in simplex often differ from those in real space. A detailed description of the D-part simplex as the sample space of compositional data and how its geometry differs from real space can be found elsewhere (Barceló-Vidal & Martín-Fernández, 2016; Egozcue, Pawlowsky-Glahn, Mateu-Figueras, & Barceló-Vidal, 2003; Pawlowsky-Glahn, Egozcue, & Tolosana-Delgado, 2015). Most commonly used statistical methods.
in epidemiological research are designed for unconstrained data in real space. Although these methods are generally robust, fundamental differences between the mathematical properties of the \textit{D-part simplex} and real space make it inappropriate and risky to use them for the analysis of untransformed compositional data. As shown by Dumuid and colleagues (2017) on a typical time-use epidemiological real-data example, such practice may lead to invalid estimates and potentially misleading conclusions.

There have been a number of different attempts to fit compositional data into other, so-called “simpler sample spaces” than the \textit{D-part simplex}, but they fail to acknowledge the true mathematical properties of compositional data (for more information about this topic, see Aitchison, 2005, pp. 111-114). One such statistical approach that has gained increasing popularity in the field of time-use epidemiology over the last couple of years is \textit{Isotemporal Substitution Modelling} proposed by Mekary and colleagues (2009). Mekary’s isotemporal substitution model is based on the \textit{Isocaloric Substitution Model 2a} presented in Willett (1998). The Mekary’s isotemporal substitution model aims to estimate the effects of re-allocating time spent in one part of a time-use composition (e.g. MVPA) to another part of the time-use composition (e.g. sedentary behaviour) on an outcome variable (e.g. BMI). The logic behind the model is simple and without careful statistical consideration it may seem reasonable. For example, to estimate the effect of replacing sedentary behaviour with MVPA (both expressed in minutes/day) on BMI, in Mekary’s isotemporal substitution model one would enter minutes spent in MVPA and the total time (i.e. the sum of minutes spent in MVPA and sedentary behaviour) as independent variables and BMI as the dependent variable in the regression analysis. The resulting regression coefficient for MVPA would represent the expected change in BMI for a 1-minute increase in time spent in MVPA, whilst the total time is kept constant. The total time can only be kept constant if the 1-minute increase of time spent in MVPA is followed by a corresponding 1-minute decrease of time spent in sedentary behaviour. Therefore, the regression coefficient for MVPA would represent the expected change in BMI when one minute of sedentary behaviour is replaced with one minute of MVPA. This logic may seem intuitive, but Mekary’s isotemporal substitution model cannot be considered an appropriate method for analysing time-use data, because it treats time-use variables as unconstrained vectors in real space instead of acknowledging their compositional properties and the \textit{D-part simplex} as their true sample space.

Statistical methods acknowledging the \textit{D-part simplex} as the sample space of \textit{D-part compositional data} started emerging nearly 40 years ago (Aitchison, 1986; Atchison & Shen, 1980; Pawlowsky-Glahn, et al., 2015), and have since been successfully implemented in several research fields, particularly in geology. When dealing with time-use data, it is recommended to use compositional data analysis to produce unbiased findings and to avoid potentially serious methodological issues caused by disrespecting geometry of the data’s true sample space (Chastin, Palarea-Albadejo, et al., 2015; Dumuid, Stanford, et al., 2017; Pedišić, 2014).

\textbf{Compositional data analysis for time-use epidemiology}

Although compositional data analysis is not new to researchers investigating health-related components of time use, its adoption has been relatively slow until recent times. In 2002, as part of a symposium entitled “Analysis of Compositional Data: Problems and Solutions” at the \textit{Annual Convention of the American Alliance for Health, Physical Education, Recreation and Dance}, Zhu, Ainsworth and Liu presented their study comparing physical activity patterns of black and white women using the compositional data analysis approach (Zhu, Ainsworth, & Liu, 2002). To the best of our knowledge, this was the first application of compositional data analysis in the field of physical activity epidemiology. In 2013, in a doctoral thesis entitled \textit{On Compositional Data Modeling and Its Biomedical Applications}, Zhang analysed relationships between age, sex, ethnicity, education level, and employment status as independent variables with the selected components of energy expenditure; namely sleeping, sitting quietly/watching TV, grooming, and sitting while eating as dependent variables, to demonstrate a possible application of compositional data analysis (Zhang, 2013). In the same year, Taylor and colleagues were the first to explore changes in physical activity over time using compositional data analysis in a longitudinal accelerometer-based study among young children (Taylor, Williams, Farmer, & Taylor, 2013). The same group later analysed associations of selected socio-demographic and lifestyle variables with \textit{SLP-SB-LPA-MVPA}, and also interdependencies between the parts of the composition (Williams, Farmer, Taylor, & Taylor, 2014). In 2014, Pedišić presented methodological rationales for the use of compositional data analysis when analysing the relationship of sleep, sedentary behaviour, LPA, and MVPA with health outcomes (Pedišić, 2014). A year later, Chastin and colleagues published the first paper exploring relationships between the time-use composition and health outcomes using compositional data analysis (Chastin, Palarea-Albadejo, et al., 2015). This was followed by a range of studies using the compositional data analysis approach to deal with sleep, sedentary behaviour, and physical activity data (Carson, et al., 2016; Dumuid, et al.,
To avoid inconsistencies in reporting in this fast developing area, it is important to use standard terminology for compositional data analysis. Herewith we therefore provide a brief overview of the commonly used methods for the analysis of compositional data and associated terms in the context of time-use epidemiology. In an $n \times D$ table including data of $n$ participants (row vectors $x_1,\ldots,x_n$ in $D$ variables (column vectors $w_1,\ldots,w_D$), the set of $D$ variables is referred to as a composition and every column vector as a part of the composition. Accordingly, in time-use epidemiology, time-use variables (e.g. amounts of time spent sleeping, in sedentary behaviour, in LPA, and in MVPA) are parts of the time-use composition.

The distance between two $D$-part compositional vectors $x = [x_1,x_2,\ldots,x_n]$ and $y = [y_1,y_2,\ldots,y_n]$ in $D$-part simplex, referred to as Aitchison or compositional distance ($d_c$), can be calculated as:

$$d_c = \sqrt{\sum_{i=1}^{D} \left( \ln \frac{x_i}{g(x)} - \ln \frac{y_i}{g(y)} \right)^2}, \quad (1)$$

where $g$ is the geometric mean. Commonly used descriptive statistics, designed for variables in real space, such as arithmetic mean and standard deviation, do not fit with the geometry of compositional data, that is, Aitchison geometry (Pawlowsky-Glahn, et al., 2015). An appropriate measure of central tendency for a 24-hour time-use composition is the centre ($g$), i.e. compositional mean, calculated as

$$g = C[g_1,g_2,\ldots,g_D], \quad (2)$$

where $g_j$ is the geometric mean of $j$-th part of the composition and $C$ is the closure operator to 24-hours. Instead of using standard deviations of individual parts, dispersion of time-use data should be described using the variation matrix ($T$) calculated as

$$T = [t_{jk}] = \begin{bmatrix}
0 & \text{var}(\ln \frac{w_1}{w_2}) & \cdots & \text{var}(\ln \frac{w_1}{w_D}) \\
\text{var}(\ln \frac{w_2}{w_1}) & 0 & \cdots & \text{var}(\ln \frac{w_2}{w_D}) \\
\vdots & \vdots & \ddots & \vdots \\
\text{var}(\ln \frac{w_D}{w_1}) & \text{var}(\ln \frac{w_D}{w_2}) & \cdots & 0
\end{bmatrix} \quad (3)$$

where $t_{jk} = \text{var}(\ln \frac{w_j}{w_k})$, i.e. the variance of log-ratios of corresponding elements of $j$-th and $k$-th parts of the time-use composition. If parts $j$ and $k$ are perfectly proportional, $t_{jk}$ will be equal to zero. Higher values of $t_{jk}$ will denote lower proportionality between parts $j$ and $k$. The elements in the main diagonal of the matrix are always equal to zero, because they show the proportionality of each part with itself. Total variance of a composition can be expressed as the sum of all entries below or above the main diagonal in the variation matrix $T$ divided by $D$, which corresponds to the average squared Aitchison distance between compositional vectors $x_1,\ldots,x_n$ and the centre. The centre minimises the total variance of compositional data, because the sum of Aitchison distances between the centre and compositional vectors $x_1,\ldots,x_n$ is minimal, which is not a property of the vector of arithmetic means. The centre may substantially differ from the vector of arithmetic means; hence, arithmetic mean should not be used as a descriptive statistic for an individual part of the time-use composition. Detailed explanations of how to interpret the centre, variation matrix, total variance, and other potentially useful descriptive statistics for time-use data can be found in Pawlowsky-Glahn et al. (2015). Basic description of compositional data can be graphically supported using ternary diagrams, geometric mean barplots, canonical variates plots, compositional biplots, and coda-dendrograms. These can be used, for example, to display the centre and scattering of individual data points, to depict the relationship between parts of the composition and their multivariate structure, or to identify outliers in the dataset (Aitchison & Greenacre, 2002; Filzmoser, Hron, & Reimann, 2012; Martin-Fernández, Daunis-I-estadella, & Mateu-Figueras, 2015; Pawlowsky-Glahn, et al., 2015). Examples of such graphical presentations of $\odot$SLP-SB-LPA-MVPA data can be found in Williams et al. (2014) and Chastin et al. (2015).

Furthermore, a variety of compositional data analysis methods for establishing associations between variables, testing between-group differences, clustering data, and reducing dimensionality are available to support common research questions and study designs in time-use epidemiology (Aitchison, 1986, 2005; Pawlowsky-Glahn, et al., 2015). When performing such analyses on compositional data, it has been recommended to follow the principle of working on coordinates with respect to an appropriately selected orthonormal basis (Mateu-Figueras, Pawlowsky-Glahn, & Egoscue, 2011). Expressing
compositional data as orthonormal coordinates, usually referred to as isometric log-ratio (ilr) coordinates, is especially useful for time-use epidemiology, as it enables the application of a whole spectrum of statistical analyses for unconstrained real data that researchers in this area are traditionally familiar with. Another advantage of this approach is that it allows including non-compositional covariates (e.g. age, sex, timing of activity, frequency of sit-stand breaks, average bout length) in the statistical model alongside the ilr coordinates. Different types of ilr coordinates can be constructed, depending on the orthonormal basis of reference. The orthonormal basis may be chosen to allow for intuitive interpretation of the transformed data and/or results of a subsequent data analysis. A row vector of ilr coordinates, \( \mathbf{z} \), with respect to a basis associated with a sequential binary partition that has been shown particularly useful in time-use epidemiology, can be calculated as:

\[
\mathbf{z} = [z_1, z_2, \ldots, z_{D-1}] = ilr(\mathbf{x}) = \left[ \frac{D-1}{D} \ln \left( \frac{x_1}{\sqrt[r]{\prod_{k=2}^{D} x_k}} \right), \frac{D-2}{D-1} \ln \left( \frac{x_2}{\sqrt[r]{\prod_{k=3}^{D} x_k}} \right), \ldots, \frac{1}{\sqrt{D}} \ln \left( \frac{x_{D-1}}{x_D} \right) \right],
\]

(3)


A time-use composition of \( D \) parts will be represented by \( D-1 \) ilr coordinates, which account for the total variance of the composition. When using the above sequential binary partition and ilr coordinate representation, all relative information about the first part of the composition (\( x_1 \)) is included in the first ilr variable (\( z_1 \)). The remaining ilr variables (\( z_2, \ldots, z_{D-1} \)) contain no information about \( x_1 \). It is required for a meaningful interpretation of a statistical analysis to represent each compositional part using a single ilr variable, the parts can be iteratively rearranged in \( D \) permutations to place each one of them as the first part.

Other ilr coordinates useful for research in time-use epidemiology can be obtained from different sequential binary partitions where parts are grouped into interpretable subsets (Egozcue & Pawlowsky-Glahn, 2005). For example, sleep and sedentary behaviour can logically be grouped into an ‘inactivity’ subset, whilst LPA and MVPA can be grouped into an ‘activity’ subset. The ilr coordinates obtained in this way are often referred to as balances.

Ilr coordinates can be useful when analysing between-group differences in health-related time-use variables using compositional MANOVA or discriminant analysis (Filzmoser, Hron, & Templ, 2012; Martin-Fernández, et al., 2015), conducting compositional factor analysis (Filzmoser, Hron, Reimann, & Garrett, 2009), clustering compositional data (Aitchison, Barceló-Vidal, Martin-Fernández, & Pawlowsky-Glahn, 2000; Martin-Fernández, Barceló-Vidal, & Pawlowsky-Glahn, 1998; Palaarea-Albaladejo, Martin-Fernández, & Soto, 2012), analysing correlations and canonical correlations between compositional variables (Filzmoser & Hron, 2009), and modelling compositional data as explanatory or outcome variables in regression analysis (Dumuid, Stanford, et al., 2017; Hron, et al., 2012; Pawlowsky-Glahn, et al., 2015). Examples of application of compositional cluster analysis on SLP-SB-LPA-MPA-VPA can be found in Dumuid et al. (2016), Dumuid, Olds, Lewis, et al. (2017), and Dumuid, Olds, Martin-Fernández, et al. (2017). A straightforward method to interpret regression coefficients when analysing compositions with a meaningful total (e.g. 24 hours/day) was recently proposed by Dumuid and colleagues (2017). The paper includes an empirical analysis of the relationship between SLP-SB-LPA-MVPA and adiposity, as an example of how the proposed method can be used in time-use epidemiology, as well as a description and a practical example of using balances to analyse the relationship between the time-use composition and health outcomes. To avoid the issues associated with Mekary’s isotemporal substitution, Dumuid and colleagues (in press) proposed a compositional isotemporal substitution model. An example of this analysis can be found in Fairclough et al. (2017), where isotemporal substitutions between parts of SLP-SB-LPA-MVPA were associated with measures of fitness and adiposity among children. Some useful guidelines for conducting compositional data analyses on health-related time-use variables can be found in a supplementary file published alongside the Chastin et al. (2015) paper.

Depending on the classification criteria, time-use data may contain rounded or trace zeros, usually
arising from the inability of a measurement tool to correctly record short bouts of time spent in a time-use component, and essential or structural zeros, when truly no time was spent in a particular time-use component. Presence of any of the two types of zeros in the dataset will preclude the calculation of ilr coordinates, because the logarithm of zero is undefined. Possible solutions and strategies for dealing with zeros can be found elsewhere (Bear & Billheimer, 2016; Martin-Fernández, Hron, Templ, Filzmoser, & Palaera-Albaladejo, 2012; Martin-Ferrández, Palaera-Albaladejo, & Olea, 2011). A careful consideration that takes into account the criteria used for time-use classification, measurement protocols, and characteristics of the study population needs to be done in every individual study, before categorising the observed zeros in one of the two abovementioned categories.

Detailed user-friendly instructions on how to analyse compositional data using R software can be found in van den Boogaart and Tolosana-Delgado (2013). Basic exploratory analyses of compositional data can be conducted using freely available software CoDaPack (Thió-Henestrosa & Comas, 2016; Thió-Henestrosa & Martin-Fernández, 2005). For more advanced analyses of compositional data, there are several R packages available, including “complmrob” (Keplinger, 2015), “compositions” (van den Boogaart, Tolosana, & Bren, 2015), “robCompositions” (Templ, Hron, & Filzmoser, 2017), and “zCompositions” (Palarea-Albaladejo & Martin-Fernandez, 2016).

Theoretical framework and future directions for research in time-use epidemiology

As a theoretical model for future studies, we propose the Framework for Viable Integrative Research in Time-Use Epidemiology (VIRTUE framework) (Figure 3). The framework incorporates key elements of the social-ecological approach (Sallis, et al., 2006; Stokols, 1992, 1996), Behavioural Epidemiology Framework (Sallis, Owen, et al., 2000), and the Activity Balance Model (Pedišić, 2014).

The VIRTUE framework incorporates research on: 1) methods in time-use epidemiology; 2) outcomes of time-use composition; 3) optimal time-use balance and its prevalence in populations; 4) determinants and correlates of time-use composition; and 5) effectiveness of time-use interventions. All these areas of research must be covered to provide a detailed understanding of the prevalence causes, and consequences of optimal time-use balance and ways to increase its prevalence at the population level. Ideally, the sequencing of studies would strictly follow the above order. However, based on the experiences from other fields of epidemiology, it is more realistic to expect that studies will be conducted simultaneously across all the areas of research.

Research area 1: Methodological research in time-use epidemiology

Methodological research should enable studies in time-use epidemiology to employ adequate measures of time use. The most common instruments for assessing time use in population studies have been time-use surveys, recalls, and diaries, and, more recently, device-based measures, such as accelerometers and multi-sensor devices. Previous studies in the fields of time use, sleep, physical activity, and sedentary behaviour provided a large knowledge base on measurement properties of such instruments (Foley, Maddison, Olds, & Ridley, 2012; Healy, et al., 2011; Helmerhorst, Brage, Warren, Besson, & Ekelund, 2012; Pedišić & Bauman, 2015; Ridley, Olds, & Hill, 2006). Nevertheless, there are still a number of open questions regarding the optimal measurement of time-use components that must be addressed by future research, particularly with regard to accelerometer-based measures and other device-based activity recognition systems (Pedišić & Bauman, 2015; Rasalinghe, Al MacHot, & Mayr, 2016). Besides, previous measurement studies in this area have treated each part of the time-use composition individually. This could have resulted in wrong
estimates of validity and reliability, because such an approach does not acknowledge the compositional properties of time-use data. Future measurement studies should, therefore, focus on: [i] development of new and improvement of existing measurement tools for assessing the health-related components of time use; and [ii] establishing the measurement properties of existing and newly developed measurement tools using compositional data analysis.

The main function of health surveillance systems, such as the Australian Health Survey (AHS), Health Survey for England (HSE), and the US National Health and Nutrition Examination Survey (NHANES), is to inform policy makers and other public health stakeholders about prevalence rates of diseases and ‘unhealthy’ behaviours in populations and their trends over time. Health surveillance systems are also one of the key sources of data for epidemiological research. Time-use epidemiologists should, therefore, advocate for appropriate measurement of the most relevant time-use variables in national health surveillance systems. Valuable data on health-related components of time use are regularly collected within a number of national time-use surveys (e.g. in Australia, UK, USA, Thailand). Unfortunately, very often these surveys do not include measures of health outcomes and the time-use data are not linked to medical records of morbidity and mortality. To support the development of time-use epidemiology, efforts should be made to broaden the scope of time-use surveys by including measures of health. The utility of time-use surveys for epidemiological research could also be enhanced by supporting between-study and cross-country harmonisation of data collection and processing protocols, and ensuring that all relevant health-related components of time use can be extracted from the collected data.

Another important area of methodological research is the development and evaluation of classification systems for deriving health-related components of time use from surveys and device-based data. Consensus on classification criteria is necessary to ensure comparability of findings across different studies. Promising development in this regard can be seen in a recent work by the Sedentary Behavior Research Network (SBRN) resulting in consensus among experts on the classification of movement and non-movement behaviours (Tremblay, et al., 2017). Similar processes may be needed to reach agreement among researchers on criteria to classify other relevant health-related components of time use. Further research is also needed on statistical methods for time-use epidemiology, in order to: [i] estimate the amount of bias in effect sizes obtained using methods for unconstrained data in previous sleep, physical activity, and sedentary behaviour studies; [ii] develop new or adapt existing compositional data analyses to support all common study designs; and [iii] demonstrate how to analyse time-use data using the already established compositional data analysis methods.

Research area 2: Outcomes of health-related time-use compositions

Findings on health outcomes of sleep, sedentary behaviour, physical activity, and other components of time use are greatly influential in shaping public health strategies. Most previous research on health outcomes of specific components of time use, such as sleep, sedentary behaviour, and physical activity, have not used adequate statistical methods to account for the compositional nature of time-use data (Pedišić, 2014). It is, therefore, vital for time-use epidemiology to re-examine potentially biased previous findings using appropriate statistical techniques.

Good health is not merely the absence of disease and frailty; it also implies complete physical, mental, and social well-being (World Health Organization, 2006). Future studies should, therefore, investigate all biologically, psychologically, and/or socially plausible relationships between time-use compositions and health. Focus should be placed on the most relevant health outcomes for which previous research has already provided indications of association with time-use components (Cappuccio, et al., 2011; Cappuccio, et al., 2010a, 2010b; de Rezende, et al., 2014; Koski, et al., 2017; Lee, et al., 2012; St George, et al., 2014; Xi, et al., 2014; Zhao, et al., 2013). It is also of public health and wider social significance to investigate how health-related time-use compositions affect: other lifestyle and behavioural characteristics such as smoking, alcohol intake, diet, and drug abuse; academic achievement; physical and cognitive skills and abilities; direct and indirect economic costs; and the environment. Evidence of these associations may help convince policy makers to invest more resources into the promotion of healthy time use.

Research in this area should determine: [i] overall relationship between time-use composition and its hypothesised outcomes; [ii] dose-response relationships between specific health-related time-use variables and their outcomes; [iii] effects of re-allocation of time between time-use components (e.g. displacement of 30 minutes of sedentary behaviour with 30 minutes of MVPA); and [iv] underlying mechanisms of the relationships between health-related time-use compositions and their outcomes.

Research area 3: Time-use composition: optimal balance, prevalence, and trends

To inform policy makers and help shape public health strategies, it is important to determine how widespread healthy and unhealthy time-use habits
are among the population. Average levels of or time spent in specific health-related time-use components, such as sleep, sedentary behaviour, and MVPA, and their trends have been extensively covered in the literature (Bin, Marshall, & Glozier, 2012; de Rezende, et al., 2016; Hallal, et al., 2012; Ng & Popkin, 2012). However, the findings have mostly been reported using arithmetic means or medians, ignoring the true statistical properties of time-use data. A measure of central tendency should be selected according to distributional properties of the data. Whilst, for example, for skewed data in real space the median is usually a more appropriate measure of central tendency than the arithmetic mean, compositional data are best represented using compositional means. Previous estimates of average levels of health-related time-use components and their respective time trends may therefore not best reflect the true ‘average’ levels in the population.

Reporting against defined public health benchmarks (e.g. prevalence of adults meeting the MVPA recommendation of 150 minutes/week) has been another widespread way of describing population-level status of specific health-related time-use components (Hallal, et al., 2012). The relevance of such prevalence rates is questionable, because the benchmarks have been developed based on dose-response relationships estimated from analyses designed for unconstrained real data. Methodologically sound evidence on the optimal balance between time spent sleeping, in sedentary behaviour, in LPA, and in MVPA is still scarce. Optimal balance between time spent in health-related time-use components is not necessarily the one that produces the greatest health or other benefits. To become a habitual time-use pattern for most individuals, it also needs to be biologically, psychologically, and socially attainable and sustainable over longer periods of time. That is the reason why research on optimal time-use balance is placed centrally in this framework. Finding optimal time-use balance will require consideration of both determinants and outcomes of health-related components of time use. It is likely that different outcomes will relate to different optimal time-use patterns, so an issue of particular interest will be how to find the ‘optimum of optima’. It may also be that there is no single optimal time-use balance, but a variety of optimal time-use patterns that produce similar overall benefits. In such cases, it might be useful to run discrete choice experiments (DCEs) to elicit preferred time-use trade-offs in the population of interest (Mangham, Hanson, & McPake, 2009; Ryan, Bate, Eastmond, & Ludbrook, 2001; Whitty, et al., 2014). Fulfilling these tasks should be set as a priority for epidemiologists, as it will allow for issuing evidence-based public health recommendations for time use.

Research using adequate analytical approaches is, therefore, needed to: [i] determine average time spent in different health-related components of time use; [ii] find the optimal time-use balance to prevent adverse health outcomes and gain other benefits; [iii] determine the prevalence of the optimal time-use balance among populations and specific population subgroups; [iv] identify the most common unhealthy time-use patterns in different populations; and [v] establish and track population-level trends in health-related time-use patterns over time.

**Research area 4: Determinants of optimal time use**

Knowing what makes some people less and others more likely to have unhealthy time-use habits is essential for designing effective public health interventions and targeting them to the population groups at greatest risk. Numerous studies have investigated determinants of health-related time-use components (Bauman, et al., 2012; Chastin, Buck, et al., 2015; Knutson, 2013; Sallis, Prochaska, & Taylor, 2000; Trost, Owen, Bauman, Sallis, & Brown, 2002). However, almost all of these studies analysed their associations with individual time-use components separately, not appreciating the compositional properties of time-use data, potentially leading to false estimates of effect sizes. If a factor affects one part of a time-use composition, it necessarily also affects one or more of the remaining parts of the composition. It is, for example, possible that a factor benefits one health-related component of time use (e.g. increases time spent in MVPA within recommended thresholds), whilst worsening some other component (e.g. decreases sleep duration below a recommended threshold). The compositional approach is, therefore, needed for a full comprehension of which factors are associated with the highest likelihood of the optimal time-use balance.

Research in this area should primarily focus on factors which have already been identified by previous research as potentially being associated with health-related time-use components, such as: socio-demographic; health; psychological; genetic; social and cultural; policy; behavioural and lifestyle; environmental; and skill/knowledge-related characteristics (Bauman, et al., 2012; Chastin, Buck, et al., 2015; Knutson, 2013; Sallis, Prochaska, et al., 2000; Trost, et al., 2002).

Research in the general population and among different population subgroups using appropriate methods of data analysis is needed to identify correlates and determinants of: [i] specific health-related components of time use; [ii] the optimal time-use balance; and [iii] the most common unhealthy time-use patterns.
Research area 5: Time-use interventions

The ultimate goal for time-use epidemiologists and other public health workers interested in promoting healthy time use is to develop and implement effective population-level time-use interventions aimed at achieving optimal time-use balance and/or preventing unhealthy time-use habits. Effectiveness of interventions to achieve adequate sleep duration, reduce time spent in sedentary behaviour, and increase time spent in physical activity has been tested in many previous studies (Heath, et al., 2012; Shrestha, et al., 2016; Thraen-Borowski, Ellingson, Meyer, & Cadmus-Bertram, 2017; Vézina-Im, Moreno, Nicklas, & Baranowski, 2017; Yoong, et al., 2016). It is impossible to change only one part of a time-use composition, without affecting at least one of the other parts of the composition. Interventions targeting a single time-use component while letting other components freely adjust themselves may lead to compensatory effects, which has been described in physical activity literature under the ActivityStat hypothesis (Gomersall, Rowlands, English, Maher, & Olds, 2013). It may be reasonable to intervene on two parts of the time-use composition, while keeping the other parts constant, i.e., a “swapping” strategy. Developing interventions to affect several or all parts of a time-use composition simultaneously is another way of acknowledging the compositional structure of time-use data. Selection between the two-component time reallocation, multi-component, or full-composition approach will depend on the time-use pattern of the targeted individuals. To achieve the optimal time-use balance, in some individuals, the first, potentially the simplest, approach might be sufficient, whilst among others the latter two approaches would be needed. Which ever of the three possible intervention approaches is evaluated in a trial, the statistical analysis of the effectiveness needs to account for all parts of the time-use composition and their compositional properties. Using statistical methods suitable for compositional data, intervention trials in time-use epidemiology should, therefore, evaluate effectiveness of informational, behavioural, social, environmental, and policy interventions aimed at improving time-use habits and/or achieving the optimal time-use balance in the populations or population groups by: [i] re-allocating time between two or more parts of a health-related time-use composition, whilst making sure the remaining parts are kept constant; or [ii] simultaneously affecting all parts of a health-related time-use composition.

Research in this area should also cover audit, analysis, and/or evaluation of health-related time-use policies. Typical policies issued to support promotion of healthy time use will include: public health guidelines for optimal time use, such as the Canadian 24-Hour Movement Guidelines for Children and Youth (Tremblay, et al., 2016); national and sub-national strategies and action plans; records documenting interventions to promote healthy time use in specific settings (e.g. school-based time-use policies, and workplace time-use policies); and other binding and non-binding governmental and non-governmental documents. Not all policies can be considered interventions, as they may not have yet been implemented (e.g. general non-documented position of a politician towards the promotion of healthy time use). Such policies may also have large impact on forming policy interventions, and should, therefore, be investigated as part of time-use epidemiology.

Concluding remarks

Nearly 70 years of valuable sleep, sedentary behaviour, physical activity, time-use, and mathematical/statistical research has led to the development of time-use epidemiology. Without findings from these research fields, it would be impossible to realise the necessity of shifting the traditional paradigm towards the compositional approach and conceptualise the new integrative field. It is likely that for many studies in the near future it will not be possible to follow the VIRTUE framework, because their existing datasets or measures do not allow for applying the proposed analytical approach. The concepts presented in this paper should not by any means discourage authors of such studies, because, even despite potential limitations, their findings may greatly contribute to the epidemiological evidence base.

The efficiency of transition to the VIRTUE framework will depend on how well are new ideas and knowledge disseminated among researchers. To support this process and future research collaborations in the field of time-use epidemiology, we herewith announce the establishment of the International Network of Time-Use Epidemiologists (INTUE). The word intue is an archaic form of the verb intuit, which stands for “to know, sense, or understand by intuition” (Merriam-Webster Inc., 2017). We are convinced that studies following the intuitive framework for research presented in this paper will contribute to developing a strong epidemiological knowledge base and help improve the health of populations.

It is estimated that 2.5%, 3.8%, and 9% of deaths are attributable to the lack of sleep, excessive sedentary behaviour, and insufficient physical activity, respectively (de Rezende, et al., 2016; Eguchi, et al., 2012; Lee, et al., 2012). It is likely that in total more deaths can be attributed to unhealthy SB-SB-LPA-MVPA than to smoking or obesity, potentially making this time-use composition the most relevant modifiable behavioural and lifestyle risk factor of our time. We hope that governments and leading health organisations will recognise enormous importance of healthy time use, and provide adequate support for future research in time-use epidemiology.
References


Pedišić, Ž. (2014). Measurement issues and poor adjustments for physical activity and sleep undermine sedentary behaviour research – The focus should shift to the balance between sleep, sedentary behaviour, standing and activity. *Kinesiology, 46*(1), 135-146.


