CHAPTER 25

Synthesis: Combining Methods for the Analysis of Vulnerability Processes Across the Life Course

André Berchtold and Paolo Ghisletta

Introduction

The analysis of vulnerability processes throughout the life course implies two main methodological difficulties. On the one hand, it requires long sequences of data that ideally should cover the entire lifespan of an individual, from birth to death (possibly, one may even go as far back as the womb if investigating perinatal determinants of vulnerability processes, such as exposure to detrimental environmental factors like lead, alcohol,

A. Berchtold (✉)
Faculty of Social and Political Sciences, Institute of Social Sciences, University of Lausanne, Lausanne, Switzerland
e-mail: andre.berchtold@unil.ch

P. Ghisletta
Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

Swiss Distance University Institute, Valais, Switzerland
e-mail: Paolo.Ghisletta@unige.ch

413

micronutrients, and various toxins). On the other hand, people in vulnerable conditions, or in situations that increase the risk of becoming vulnerable, are often more difficult to assess in empirical research: They either belong to hard-to-reach subpopulations or suffer from conditions that limit the chances of collecting complete data about them. Taken together, these two characteristics can require both the use of unconventional data collection methods and the development of advanced statistical analysis procedures that are able to extract a maximum amount of information from the data. Clearly, both data collection and data analysis methods need to be motivated primarily by theory rather than merely being the result of practical or empirical limitations stemming from the difficulties mentioned above. Substantive questions need to remain the main instigator that defines how data are to be collected and analysed. Otherwise, the interpretation of any result will either be impenetrable or too remote from the previous and extant substantive knowledge to be of any use to the researcher.

In this synthesis, we look back at more than ten years of methodological innovations developed within the LIVES research program and aiming for a better understanding of the development of vulnerability processes throughout the life course. Regarding data collection, although prospective surveys often remain the gold standard for obtaining longitudinal data, retrospective data collected through specially adapted tools (see chapter by Morselli & Berchtold), such as life calendars, have become precious companions, as they allow for the rapid completion of long and valid data sequences. Moreover, alternative sampling schemes, based, for instance, on networks (Perrenoud et al.), are able to reach people who are unlikely to be included in traditional probabilistic samples, thereby increasing the representativeness of the sample by including participants who otherwise could not have been assessed. Such approaches also require the development and use of more global data collection modes that combine qualitative and quantitative data, thereby allowing participants to respond in the manner best suited to their specific situations and capacities. However, the increasing complexity of data collection procedures also implies additional risks regarding the comparability of data to existing knowledge and the psychometric quality of measurement. For example, what can be gained, on the one hand, through the simultaneous use of several data collection modes can be lost, on the other hand, because data collected through one specific mode can prove to be not fully comparable with that obtained through another, whether in terms of margin of error,

number of missing data, or representativeness. Special care thus needs to be taken when implementing such advanced data collection schemes to avoid jeopardising the whole research enterprise.

What is true regarding data collection is also a requirement for accurate data analysis: From a substantive perspective, relying on a single approach to analyse complex data is rarely sufficient to deepen the theoretical understanding of a particular vulnerability topic. Indeed, it is often essential to combine the strengths of different analytical tools. One approach is to rely on mixed methods to combine qualitative and quantitative information (Bryman, 2008; Joly-Burra et al., 2020), but even if only quantitative approaches are implemented, it is sensible, for instance, to associate sequence analysis with survival analysis, the aim being to study the occurrence of specific events (survival analysis) without losing sight of an individual's prior trajectories on a categorical variable over their life course (sequence analysis). This approach is also possible if prior trajectories are represented by continuous variables so that traditional longitudinal methods, such as linear mixed-effects models, can be conjointly estimated with survival analyses.

In this chapter, we discuss the new perspectives in terms of data collection and data analysis offered by the different methodological tools developed within LIVES.

Longitudinal Data Collection Among Vulnerable Populations

Some decades ago, it was possible for a social scientist to draw data from a cross-sectional survey, perform analyses on life course issues, and publish the results. Clearly, scholars were aware that they were approximating within-person processes by examining between-person data, but costs for repeated data collections were prohibitive, and methods to handle statistical artefacts of such data (such as dependence) were of the sole domain of methodologists and statisticians; hence, a restricted population of scholars was not necessarily confronted with substantive research themes. However, following the development of new theories in life course research and the spread of appropriate methods for repeated-measures data to large audiences of substantively motivated scholars, the situation has evolved, and we are now in a world in which longitudinal analyses, and hence data, are the rule for life course research. For instance, in their definition of vulnerability, Spini et al. (2017) stressed that in addition to being a

multidimensional and multilevel concept, vulnerability also develops and evolves over time. This is not to say that it is no longer possible to conduct sound cross-sectional research, but it is now established that the present situation of an individual is strongly related and/or influenced by her or his past history. Therefore, to understand a given individual's situation, scholars need to consider that individual's life as a whole and, consequently, use longitudinal data and longitudinal analyses. Note, however, that we are not completely rejecting the use of cross-sectional data. Indeed, cross-sectional methods remain highly valuable when investigating life-course vulnerability processes, given that any longitudinal study starts out as a cross-sectional study. Particularly in novel scholarly terrain, a cross-sectional study remains a feasible and invaluable exploration tool that can shape the beginning of a successful longitudinal research enterprise.

Generally, collecting longitudinal data is notoriously more difficult than collecting cross-sectional data because it requires multiple data collection points, hence the requirement to motivate people to answer to multiple assessments over time. The first, and at times most serious, drawback of a longitudinal study is attrition, i.e., the fact that as time unfolds, increasingly fewer initial participants continue assessment. Many strategies have been developed to reduce attrition, including the possibility of answering through various data collection modes (Voorpostel et al., 2021) and different kinds of incentives offered to the participants (Stähli & Joye, 2016). Particular care needs to be taken to ensure that the returning participants adequately represent the initial sample well; otherwise, generalisability is reduced. However, even if such strategies are effective against attrition, they cannot reduce the time required to collect sufficiently long datasets. Accelerated longitudinal designs (Bell, 1953; Galbraith et al., 2017), which consist of assessing multiple cohorts for a shorter interval but with temporal overlap among cohorts, are an extremely interesting alternative, but such data are not compatible with all analysis methods (i.e., such data inherently require a multiple-group approach). Another approach is to collect data retrospectively, whereby participants reconstruct their past life course, but this implies other issues related mainly to the difficulty of recalling old memories and accurately reporting them. The use of a life history calendar, a tool especially designed to enhance memory recall, can mitigate many such difficulties to allow the reconstruction of valid and reliable past trajectories (Belli, 1998).

Difficulties in investigating usual general population samples become even more complex when researchers investigate minority, hard-to-reach,

or vulnerable populations. In addition to those previously mentioned, investigating such populations implies additional challenges. Principal among these is the problem of how to contact them, given that they are generally invisible either because they are not clearly distinguishable from other subpopulations or because they live on the margins of society, with very limited or no administrative contact. Therefore, imaginative ways of not only collecting data but also contacting people are required.

Three chapters of this book tackle the problem of data collection. The first discusses the use of life calendars for collecting retrospective data (Morselli & Berchtold). The second examines alternative sampling schemes that use a network of respondents to increase sample size and diversity and allow the investigation of hard-to-reach individuals (Perrenoud et al.). The third reviews the benefits but also the difficulties of offering different answering modes to survey respondents (Roberts & Voorpostel). Taken together, the conclusions of these three chapters show promise for longitudinal data collection: New methods are available, and they work. However, they also imply additional difficulties in generating a representative sample and obtaining accurate information. More work is clearly required to further explore the properties of these methods and to validate them in additional research practice. We may add to these challenges the increasing number of polls (mostly conducted by the media and too often with low scientific standards), which decrease the likelihood of finding respondents to academic surveys (Beullens et al., 2018), and the increasing use of the internet to collect information, which increases the difficulty of evaluating data quality (Benfield & Szlemko, 2006).

In this context, we must ask ourselves what we want or need to focus on. Data-scraping tools (Marres & Weltevrede, 2013), which allow the collection of very large ('big data') information bases from the internet, are now just a mouse click away through the web and popular social media. In this sense, the internet increasingly resembles a quasi-infinite data repository. However, is it an appropriate way to collect longitudinal data for scientific purposes? Is it still possible to obtain a representative sample through this approach? What biases are produced or increased? Is it ethical to collect and link data without the active knowledge of the persons concerned? All these questions are still open, and we as scholars must remain aware that technical progress must not take precedence over scientific rigor and ethics.

STRENGTHENING AND COMBINING STATISTICAL ANALYSIS TOOLS

Relying on a single statistical index in isolation has rarely been a good and sufficient analytical practice. Even one of the simplest statistics, the mean, ought to be combined with a measure of the dispersion, such as the standard deviation, to adequately represent a variable (note that, depending on the underlying distribution, many alternative indices of central tendency and variability may be preferable (Tukey, 1977; Weisberg, 1992)). This reasoning can be easily extended to much more complex analyses. The life course of an individual is made of a succession of events, or trajectories of varying magnitude, whether important or not, that shape his or her future life. These past events and trajectories deserve a closer examination to elucidate one's life course. For instance, while most individuals will experience diverse health problems during their life, what are the determinants of suffering from one specific disease after the age of 50? Thus, while from a holistic perspective an individual's life trajectory must be taken as a whole, the successive events that shape that trajectory are also of interest.

Two chapters in this book tackle the study of past sequences/trajectories as determinants of upcoming events: The first primarily considers the case of categorical data that define sequences of events (Studer et al.), and the second discusses trajectories on continuous data and their relation to upcoming (often detrimental) events (Joly-Burra et al.). As shown in both chapters, the key to the sound analysis of complex life course data is the ability not only to apply different analyses successively but also to combine them into a new, integrated methodological approach. Studer et al. described the combined use of sequence analysis and event history analysis. When the objective is to study what can lead to the occurrence (or not) of a specific event, then sequence analysis can be used first to identify frequently occurring life trajectories, and then event history analysis can use this information to predict the risk of occurrence of the event of interest. The reciprocal is also possible: The interest can be in how individuals cope with the occurrence of a specific event, e.g., how they recover from a stressor, but without forgetting that the resources accumulated before the event of interest can be helpful in coping after its occurrence. Methodologically, this approach implies identifying what can happen after the event through sequence analysis and then applying an event history model to understand what triggers the occurrence of each of the possible trajectories.

Another integrated approach is described in Joly-Burra et al. Here, the idea is to simultaneously estimate two submodels, the first for the whole longitudinal trajectory as represented by continuous data and the second to study the risk of occurrence of a specific event. These two parts are then linked by a joint model, which simultaneously estimates all parameters of both submodels. Thus, both the longitudinal and the event history submodels make up the final model, wherein the parameters of one submodel are necessarily influenced during estimation by those of the other. The methods described in Studer et al. and in Joly-Burra et al. are promising given the high level of statistical integration they propose. We urge any life-course scholar who is particularly interested in vulnerability processes to assimilate these two chapters in the hope that these integrative methods may serve their own research agenda well.

Our retrospective look at methodological contributions within LIVES over the past ten years highlights the investigation of other approaches that combine statistical methods to extract more information from data. In a typical sequence analysis approach, dissimilarities between sequences are computed to create a typology that greatly reduces the complexity of the set of individual components. This typology may then be used as an independent variable in subsequent analyses, as shown by Ritschard and Studer (2018). Alternatively, Bolano and Studer (2020) proposed adopting another approach: Instead of reducing the number of different individual trajectories, and hence their complexity, they automatically extract a very large set of indicators representing the key information (timing, sequencing, and duration) of all trajectories. Then, machine learning algorithms are used to establish potential relations between these indicators and an event of interest or to identify a minimal set of indicators that are able to accurately predict this same event.

The different statistical methods mentioned above all consider the life course of an individual in its entirety, but debate remains ongoing between supporters of this vision and scholars who prefer to consider and analyse the underlying processes that generate the sequences themselves. Using a Markovian-type model, the latter approach was applied, for instance, by Berchtold et al. (2018) to study the development of somatic complaints among adolescents and by Bolano et al. (2019) to study the process of change in functional health among nursing home residents. Furthermore, Bolano and Berchtold (2021) showed that these two visions, life trajectory as a whole vs. generating process, can—and should—be regarded as complementary rather than competitive. This integrative approach entails

a first step in which a sequence analysis is performed as usual to reduce the complexity of a dataset, and a second step in which a Markovian approach is applied independently to each subset to gain more information on the specificities of the trajectories associated with each group.

OTHER DEVELOPMENTS

Two additional issues ought to be considered for an accurate analysis of life course trajectories. First, a person's life cannot usually be summarised by any single trajectory, such as a set of occupational or family situations. Thus, the combination of trajectories from many different life domains, such as health, family, social participation and occupation, reveals the richness and complexity of a person's life. To analyse these trajectories simultaneously, we can use an extension of sequence analysis called multichannel analysis (Gauthier et al., 2010; Piccarreta, 2017). Similarly, the analysis of reciprocal dynamical influences between two or more life trajectories is also a promising tool to further the understanding of vulnerability processes (as exemplified by Aichele & Ghisletta, 2019, where the authors analysed over 100,000 individual trajectories of memory performance and of depressive symptoms, showing that the former influences changes in the latter, whereas the reciprocal effect did not emerge). In the future, it will also be of high interest to adapt the different approaches discussed in this article to the case of multichannel data, which will undoubtedly require additional mathematical developments.

Second, we simply cannot omit the inevitable topic of missing data. Regardless of the data collection method, missing data occur in almost all datasets, thereby complicating the computation of reliable statistical results when they do not simply prevent them from being obtained. As demonstrated in Berchtold (2019), even if the impact of missing data on statistical analyses is known, as are the possible remedies, most social sciences publications still do not appear to have seriously integrated these concerns. Life trajectories are not an exception, and missing data have been identified as one of the most important remaining issues to be considered in sequence analysis (Ritschard & Studer, 2018). Moreover, even if different authors have already proposed possible solutions (Gabadinho & Ritschard, 2016; Halpin, 2016), additional developments are still required to fully consider the specificities of longitudinal data (Berchtold & Surís, 2017).

Conclusion

Vulnerable populations are especially prone to being represented by low-quality data and hence low-quality statistical results because they are more likely to be hard to reach, hide their real situation, evolve quickly, and drop out of a longitudinal study because the occurrence of a life event is an additional risk of leaving a survey (Ellard-Gray et al. 2015; Kaeser, 2016). Such challenges may lead to the data minimising the possible negative, or at times positive, impact of such events. For instance, when considering representative data from a whole population and estimating a typology from such data, it is quite common to obtain, among well-defined groups, a highly heterogeneous group that is very difficult to analyse and interpret (e.g., Taushanov & Berchtold, 2018; Taushanov & Ghisletta, 2020). Such a group may consist of vulnerable individuals, but relying on a single classical clustering approach is clearly not sufficient to understand the dynamics at work in this group.

Even if their goals and their models are different, all the chapters included in the methodological section of this book lead to two main conclusions: (1) The very high complexity of life course trajectories is difficult to grasp through a single analytical approach because most tools are extremely specialised in recovering a single aspect of the data, rather than the overall multifaceted complexity; and (2) by combining two or more approaches, it is possible to extract much more information from the data, leading to a more comprehensive understanding of both the data and of the different processes at work during a life course. Such understanding is especially critical in the case of vulnerability processes because they tend to evolve across an individual's entire life course, such as through the accumulation of resources over several years that will prove useful in a later phase of his or her life to cope with adverse stressful events. A combination of longitudinal and point approaches is therefore required to consider all this information and to highlight the links among an individual's past trajectories and sequences, present situation, and possible futures.

References

Aichele, S., & Ghisletta, P. (2019). Memory deficits precede increases in depressive symptoms in later adulthood. *The Journals of Gerontology: Series B, 74*(6), 943–953. https://doi.org/10.1093/geronb/gbx183

Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development, 24*(2), 145–152. https://doi.org/10.2307/1126345

Belli, R. F. (1998). The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory, 6*(4), 383–406.

Benfield, J. A., & Szlemko, W. J. (2006). Internet-based data collection: Promises and realities. *Journal of Research Practice, 2*(2), 1–15.

Berchtold, A. (2019). Treatment and reporting of item-level missing data in social science research. *International Journal of Social Research Methodology, 22*, 431–439. https://doi.org/10.1080/13645579.2018.1563978

Berchtold, A., & Surís, J.-C. (2017). Imputation of repeatedly observed multinomial variables in longitudinal surveys. *Communications in Statistics – Simulation and Computation., 46*, 3267–3283. https://doi.org/10.1080/0361091 8.2015.1082588

Berchtold, A., Surís, J.-C., Meyer, T., & Taushanov, Z. (2018). Development of somatic complaints among adolescents and young adults in Switzerland. *Swiss Journal of Sociology, 44*, 239–257.

Beullens, K., Loosveldt, G., Vandenplas, C., & Stoop, I. (2018). *Response rates in the European social survey: Increasing, decreasing, or a matter of fieldwork efforts?* https://doi.org/10.13094/SMIF-2018-00003

Bolano, D., & Berchtold, A. (2021). The analysis of inequality in life trajectories: An integration of two approaches. In *Rutledge International Handbook on Inequalities and the Life Course*. In Press.

Bolano, D., Berchtold, A., & Bürge, E. (2019). The heterogeneity of disability trajectories in later life: Dynamics of activities of daily living performance among nursing home residents. *Journal of Aging and Health, 31*, 1315–1336. https://doi.org/10.1177/0898264318776071

Bolano, D., & Studer, M. (2020). The link between previous life trajectories and a later life outcome: A feature selection approach. *LIVES Working Papers, 82.* https://doi.org/10.12682/LIVES.2296-1658.2020.82

Bryman, A. (2008). Why do researchers integrate/combine/mesh/blend/mix/ merge/fuse quantitative and qualitative research? In M. M. Bergman (Ed.), *Advances in mixed methods research* (pp. 87–100). Sage.

Ellard-Gray, A., Jeffrey, N. K., Choubak, M., & Crann, A. E. (2015). Finding the hidden participant: Solutions for recruiting hidden, hard-to-reach, and vulnerable populations. *International Journal of Qualitative Methods, 14.* https://doi.org/10.1177/1609406915621420

Gabadinho, A., & Ritschard, G. (2016). Analysing state sequences with probabilistic suffix trees: The PST R library. *Journal of Statistical Software, 72*, 1–39.

Galbraith, S., Bowden, J., & Mander, A. (2017). Accelerated longitudinal designs: An overview of modelling, power, costs and handling missing data. *Statistical Methods in Medical Research, 26*, 374–398.

Gauthier, J.-A., Widmer, E. D., Bucher, P., & Notredame, C. (2010). Multichannel sequence analysis applied to social science data. *Sociological Methodology, 40*, 1–38.

Halpin, B. (2016). Multiple imputation for categorical time series. *The Stata Journal*. https://doi.org/10.1177/1536867X1601600303

Joly-Burra, E., Van der Linden, M., & Ghisletta, P. (2020). A mixed-method study on strategies in everyday personal goals among community-dwelling older adults. *Gerontology, 1–10*. https://doi.org/10.1159/000508824

Kaeser, L. (2016). Adapting quantitative survey procedures: The price for assessing vulnerability? lessons from a large-scale survey on aging and migration in Switzerland. In M. Oris, C. Roberts, D. Joye, & M. Ernst Stähli (Eds.), *Surveying human vulnerabilities across the life course* (pp. 65–85). Life Course Research and Social Policies, vol 3. Springer. https://doi.org/10.1007/978-3-319-24157-9_3

Marres, N., & Weltevrede, E. (2013). Scraping the social? *Journal of Cultural Economy, 6*, 313–335. https://doi.org/10.1080/17530350.2013.772070

Piccarreta, R. (2017). Joint sequence analysis: Association and clustering. *Sociological Methods & Research, 46*, 252–287.

Ritschard, G., & Studer, M. (2018). Sequence analysis: Where are we, where are we going? In G. Ritschard & M. Studer (Eds.), *Sequence analysis and related approaches* (pp. 1–11). Life Course Research and Social Policies, vol 12. Springer.

Spini, D., Bernardi, L., & Oris, M. (2017). Toward a life course framework of vulnerability. *Research in Human Development, 14*(1), 5–25.

Stähli, E., & Joye, D. (2016). Incentives as a possible measure to increase response rates. In C. Wolf, D. Joye, T. W. Smith, & Y. C. Fu (Eds.), *The SAGE handbook of survey methodological research* (pp. 425–440). Sage.

Taushanov, Z., & Berchtold, A. (2018). Markovian-based clustering of internet addiction trajectories. In G. Ritschard & M. Studer (Eds.), *Sequence analysis and related approaches* (pp. 203–222). Life Course Research and Social Policies, vol 12. Springer. https://doi.org/10.1007/978-3-319-95420-2_12

Taushanov, Z., & Ghisletta, P. (2020). The use of a hidden mixture transition distribution model in clustering few but long continuous sequences: An illustration with cognitive skills data. *Symmetry, 12*(10), 1618. https://doi.org/10.3390/sym12101618

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Publishing Company.

Voorpostel, M., Lipps, O., & Roberts, C. (2021). Mixing modes in household panel surveys: Recent developments and new findings. In P. Lynn (Ed.), *Advances in Longitudinal Survey Methodology*. Wiley.

Weisberg, H. F. (1992). *Central tendency and variability*. Sage.