



# Late Bloomer Offending Patterns: Towards a Harder Empirical Definition—Commentary on Matsuda et al. (2022)

Torbjørn Skardhamar<sup>1</sup>

Received: 21 August 2022 / Revised: 25 October 2022 / Accepted: 12 November 2022  
© The Author(s) 2022

## Abstract

Matsuda et al. (2022) show that late bloomers also exist in self-report data. They use a “dynamic” definition of late bloomer offending pattern as they have concerns with definitions based on arbitrary age cutoff points. But the approach does not make the groups less arbitrary than hard definitions. In this commentary, I argue that a hard definition has important advantages and demonstrates an alternative using a hard definition while exploring heterogeneity within the late onset group.

In a recent study, Matsuda et al. (2022), hereafter MTLK, provide new empirical evidence of the existence of late blooming offending patterns, which is understood as persons who escalate offending only after the normative peak. The main contribution by MTLK is that they use self-report data well into adulthood and thus address the concern that earlier findings using official data might be an artefact of measurements. I generally agree with MTLK’s arguments and welcome their empirical findings and conclusion. However, their “dynamic” definition of groups has some important limitations, and I believe that their findings would have been even more convincing not using a dynamic definition.

MTLK argue that a “static” definition of late bloomers with an age cutoff point would be arbitrary as well as not in line offending as “a developmental process that unfolds continuously over time” (p. 126). Moreover, a cutoff point would be overly inclusive and lead to too heterogeneous classifications. To avoid these problems, they define late bloomers based on latent class trajectory modelling (LCTM), which shows a group that escalates late in adolescence. While this might be fine, LCTM does not solve the issues of arbitrary classifications and being too inclusive.

---

This work is submitted as a commentary article.

---

✉ Torbjørn Skardhamar  
torbskar@sosgeo.uio.no

<sup>1</sup> Department of Sociology and Human Geography, University of Oslo, Oslo, Norway

The “dynamic” definition of late bloomers has two other main limitations. First, it is not quite clear who are regarded as late bloomers in this sense. The group might include persons who have been offending at a low level for some time while escalating further in early adulthood but might also include persons who have an onset only in adulthood. Relatedly, it is not entirely clear how much more offending it takes to be considered escalating. The late bloomer group might even include persons who do not bloom much at all.

For a given hypothetical pattern, it is not given how it should be classified. Classifications are based on calculation of posterior probabilities, and since these probabilities rely on the estimated model, the definition is not easily applicable to use on different dataset.

The second limitation is therefore that the study is hard to replicate on new data. The classifications are based on the specific model, and that model is dependent on the data used. Groups identified using similar methods on other datasets are not likely to capture quite the same patterns. Results from LCTM are generally hard to compare across datasets (see Skardhamar, 2009, p 872).

My concerns do not invalidate MTLK’s results nor conclusion. The point is just that their empirical approach using LCTM does not really remove arbitrariness from the classification, and we do not know from the presented results if the group is defined too inclusive or not. While MTLK’s reasoning is nuanced and reasonable, the empirical definition based on LCTM does not have any real advantages over a more clear-cut definition.

It might be possible to mitigate these limitations within their empirical approach by more fully exploring within-group heterogeneity and providing more descriptive statistics by group. It would give the reader a better understanding of what patterns are represented in the latent classes. Another possibility is to adopt a hard definition.

## Towards a Hard Definition

MTLK conceptually define late bloomers as follows: “We use the term late bloomer to refer to individuals whose trajectories of offending emerge and escalate only after the age normative peak, from approximately age 17 onward” (page 126). Thus, there are two empirical components: *emerge* and *escalate*. Each of these needs to be operationalized.

MTLK indicate that age 17 is the normative peak, so “emerging” could mean onset after this age, using a cutoff point. “Escalation” presumably means that they should commit substantial offending after onset.

A tentative suggestion is to start with late onset, which MTLK indicate could be after age 17. Escalation is more difficult to define. Early studies of criminal careers suggested that more than about five offences indicate persistent offenders (e.g. Blumstein et al., 1985), so I suggest requiring more than 5 subsequent offences.

Intuitively, I think “blooming” indicates some time span, so I suggest also requiring the offences to be spread out over at least two separate calendar years. While this definition is clearly somewhat arbitrary, it captures both late onset and escalation.

Another possibility is to do an exploratory analysis of continuation after the onset. To this end, the empirical methods used to summarize subsequent offending are of less importance, but a frequency table would be a good start. Another possibility is to use LCTM with time since onset as the running variable. One can do both.

## An Empirical Demonstration

I use data from the Norwegian police between 1992 and 2020. The data include all solved cases where the perpetrator is found, regardless of further outcomes (Lyngstad & Skardhamar, 2011). These offenders have received a formal decision so they are not merely suspects. The data include all types of offences regulated within the Criminal Act, which exclude offences related to, e.g. traffic, the environment, work-environment, and medical regulations (except drug-related crimes). The data is organized as a person-year file with yearly counts of offences.

I restrict the analysis to one single birth cohort, born 1982 and resident at the beginning of the period, but excludes those immigrating later. These are followed from the age of 10 to 38. I use a cutoff point after age 17 to define late onset but restrict to those having onset by age 33 to allow at least 5 years of follow-up after initial onset. Of this cohort, 10,595 persons had at least one offence by 2020, which is the initial sample to be analysed.

## Trajectory Model

The exploratory analysis uses LCTM on the subsample of late onsets, but restrict to those with at least one subsequent offending. The time-related variable is years since onset, where onset is time=0, and it increases by 1 each year thereafter. Since the onset is at different ages, the data has the structure of an unbalanced panel. The estimation treats these observations as missing at random although that is probably not quite true. In our case, those who do not offend due to any kind of censoring will be classified to a low-level or no-offending group, but since the interest is in the late bloomers, that is of less concern as it implies a conservative estimate.

Francis et al. (2016) find that it is better to use B-splines for the time-related variable rather than using polynomials, although not much substantive differences. Available software makes this straight-forward using, e.g. the packages *flexmix* (Leisch, 2004) and *splines* in *R* (R Core Team, 2022), so I take that approach.

Given the outcome variable is counts, the model is specified as a Poisson regression model. Searching for the number of groups is done in the standard way of comparing models based on relative fit criterion, BIC, and supplemented with substantive consideration if needed (Nagin 2005).

## Results

Figure 1 shows the cumulate proportion having committed at least one offence by age. By age 17, about 50% of the offenders have had their onset, 75% by age 25. The curve continues to increase throughout, giving a clear indication that onset occurs well into adulthood. This finding also suggests that cutoff at age 17 might be a bit early in the Norwegian context.

There is a total of 5339 who was charged for the first time after age 17, and characteristics are summarized in Table 1, grouped by their total number of offences committed. For the two first columns, offences in the year of onset are included; thus, there are no persons in the first row for zero offences. The last two columns exclude offences in the year of onset and thus denoted recidivists for number of persons and recidivism for the subsequent number of offences. The 5339 persons who had an onset after age 17 are responsible for a total of 25,743 offences. The small group of 512 persons with more than 10 offences is responsible for 54 percent of the offences.

However, 58 percent were not charged again. Thus, the majority of late onsets do not display any escalation. The remaining 2550 persons did recidivate but to

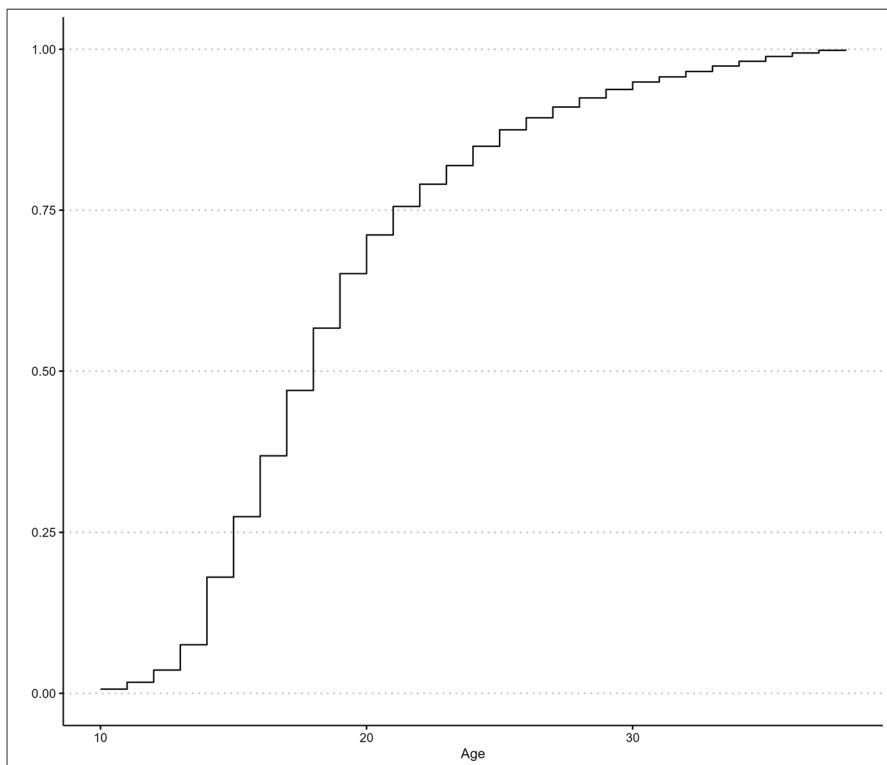


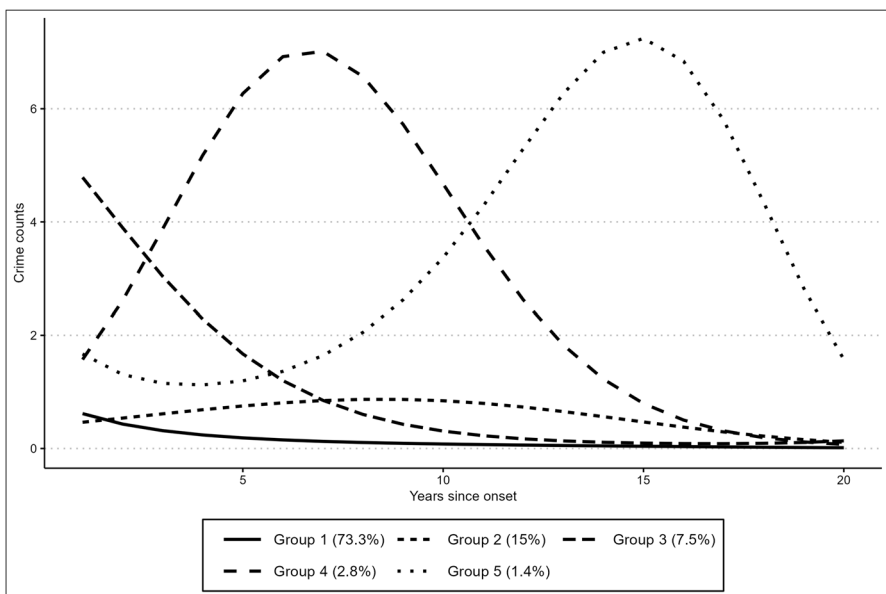
Fig. 1 Age of onset for cohort born 1982, from age 10 to 40

**Table 1** Offending frequency after age 17 until 38 for late onset. Percent of total

Number of offences per person	Persons, including onset	Offences, including onset	Recidivists (excluding onset)	Recidivism (excluding onset)	Hard definition of late bloomers (percent of group)
0			57.9		-
1–2	62.8	16.7	19.1	8.1	-
3–5	18.1	14.2	9.3	11.1	-
6–10	9.5	14.8	6.6	15.5	65.8
11–30	7.1	25.0	5.1	27.8	71.2
>30	2.5	29.3	2.0	37.5	82.0
N	5,339	25,743	5,339	17,081	712

a varying degree. If sticking to the suggested hard definition, 712 persons would be defined as late bloomers, but not everybody in the high-frequency offender groups as they apparently offended within a very limited time period.

For the exploratory analysis, LCTM are fitted to the period following onset, but excluding the non-recidivists. The year of onset is not included in the estimation, making this a description of recidivism trajectories. Figure 2 shows the chosen 5-classes model. There is no trajectory that escalates throughout; thus, any “blooming” seems to be relatively short-lived.



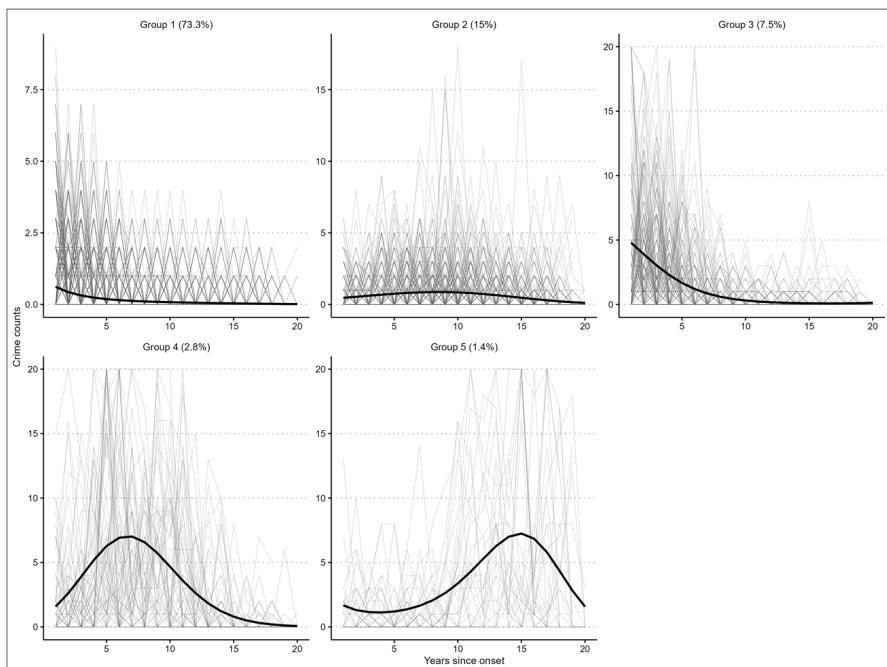
**Fig. 2** Results from latent class trajectory model by time since onset. Birth cohort 1982 with onset after age 17

To what extent these groups fit the definition of late bloomers depend on how much escalation is required. Seventy-three percent were classified into a low-level trajectory group, and another 15 percent belongs to group 2, which offend only at a slightly higher rate. These two groups do not seem consistent with the idea of late bloomers. Group 3 (7.5 percent) is desisting but from a higher rate. Maybe only trajectory groups 4 and 5 should be considered late bloomers? If so, the late bloomer group is about 4.5 percent of those with a late onset.

While the group-average trajectories are informative, there might be considerable within-group variation in offending. In Fig. 3, individual-level trajectories (grey lines) are plotted together with the group averages (black lines). The y-axis is cropped to maximum 20 offences a year to allow clearer presentations of the results. The main message from these graphs is that there is much within-group variation. The group averages represent both offending frequency and when those offences occur, while individuals' trajectories deviate considerably from the averages.

Table 2 gives some further descriptives, including how many in each group who fits the hard definition suggested above. Since groups 4 and 5 offend at a high rate, all are also late bloomers by the hard definition. But practically all in groups 2 and 3 are late bloomers by the hard definition as well. In group 1, 10 percent are late bloomer by the hard definition.

For group 1, the median number of offences is only 2, and the third quartile is 4 offences but mostly committed in a single year. Thus, the low-rate trajectory mainly



**Fig. 3** Results from latent class trajectory model by time since onset after age 17. Individual trajectories (grey) and group means (black)

**Table 2** Descriptive statistics of trajectory groups

Trajectory groups	Number of persons	Hard definition of late bloomers (percent of group)	Number of offences median	Number of offences 3rd quartile	Median number of years	Median age of onset
Group 1 (73.3%)	1675	9.8	2	4	1	20
Group 2 (15%)	316	92.4	9	13.2	5	19
Group 3 (7.5%)	167	98.2	17	25	5	19
Group 4 (2.8%)	62	100	49	63.2	9	19.5
Group 5 (1.4%)	30	100	49	60.8	11	19

reflects which years these offences are committed. For group 2, the median number of offences is 9 and spread out over a median of 5 years. Group 3 offend at a higher rate (median 17) but concentrated in the immediate years after the onset. Groups 4 and 5 offend at similar levels and spread out over more years, either with concentration early or late in the observation period. This result would be relevant to further refining the hard definition suitable for comparisons across studies.

## Conclusion

Matsuda et al. (2022) have provided a highly valuable analysis showing that there are late bloomers also in self-report data. This is particularly relevant in light of theoretical positions claiming that there is no true adult onset as they probably had an earlier unreported onset, which is not plausible for MTLK's results. I have no objections to this main finding.

However, I argue that their results would be even more convincing and useful by using a hard definition. A hard definition has a clearer interpretation and can be replicated across settings, which is a major advantage. While it is hard to avoid some arbitrary choices in such a definition, one can refine by using more than just a single indicator such as a cutoff point for age of onset. Escalation is harder to define, and that deserves further attention. While frequency and spread over time is good start, LCTM can be well suited for empirical exploration at this stage. It would be interesting to see similar analysis done with self-report data and to what extent that would affect MTLK's results.

A major advantage of a hard definition is that it requires the researchers to make an unambiguous theoretical commitment through specifying empirical consequences of the theoretical argument. However, MTLK do *not* propose a theory of late bloomers, and developing a precise definition is not really their responsibility. Their findings are particularly relevant for theories that reject that late onset exists in a meaningful way, and the definition of late bloomers should be derived from *those* theories. It is the responsibility of those advocating a theory to be sufficiently precise to make it, at least in principle, possible to refute. If that is not possible, perhaps those claims should not be taken too seriously.

**Acknowledgements** This work was supported by the Norwegian Research Council, Grant number 301883.

The authors have no relevant financial or non-financial interests to disclose.

Data has been provided from the Statistics Norway, reference no. 21/3982. Reproducible script is available at <https://osf.io/pbkyw/>.

The Data Protection Impact Assessment (DPIA) has been created in agreement with the Data Protection Official at the University of Oslo and filed with the Norwegian Centre for Research Data, project no 436491.

**Funding** Open access funding provided by University of Oslo (incl Oslo University Hospital)

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Blumstein, A., Farrington, D. P., & Moitra, S. (1985). Delinquency careers: Innocents, desisters, and persisters. In M. Tonry & N. Morris (Eds.), *Crime and Justice: A Review of Research*, 6 (pp. 187–220). University of Chicago Press.
- Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, 11(8), 1–18.
- Lyngstad, T. H., & Skardhamar, T. (2011). Nordic register data and their untapped potential for criminological knowledge. *Crime and Justice*, 40, 613–645.
- Matsuda, M., Thornberry, T. P., Loughran, T. A., & Krohn, M. D. (2022). Are late bloomers real? Identification and comparison of late-onset offending patterns from ages 14–40. *Journal of Developmental and Life-Course Criminology*, 8(2), 124–150.
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org>.
- Skardhamar, T. (2009). Reconsidering the theory of adolescent-limited and life-course persistent anti-social behaviour. *British Journal of Criminology*, 49(6), 863–878.

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