

# Co-Creation and Advanced Analytics in Register-Based Analysis of Families with Children

Marika Helkiö<sup>a,1</sup>, Anna Kuromaa<sup>a</sup>, Minna Stolt<sup>a,b</sup> and Anne Kuusisto<sup>a</sup>

<sup>a</sup>Wellbeing Services County of Satakunta, Finland

<sup>b</sup>Department of Nursing Science, University of Turku

ORCID ID: Marika Helkiö <https://orcid.org/0009-0005-2657-1981>; Anna Kuromaa <https://orcid.org/0009-0004-3388-3943>; Minna Stolt <https://orcid.org/0000-0002-1845-9800>; Anne Kuusisto <https://orcid.org/0000-0002-9950-1985>

**Abstract.** The purpose of this paper is to describe the role of co-creation in the analysis of family wellbeing data (n=22,359) concerning families with children. The target population consists of families with children from four Family Centre areas within a single wellbeing services county, based on a complete population sample. Co-creation introduced diverse expertise into the analytical process — encompassing both methodological competence and knowledge of family services — enabling the results to better support professionals working with families with children in the region.

**Keywords.** advanced analytics, artificial intelligence, k-means clustering, registry, co-creation, Families with Children, wellbeing

## 1. Introduction

The examination of families with children (hereinafter “families) and their wellbeing requires comprehensive and reliable data from multiple perspectives [1, 2]. In this paper, the wellbeing of families was examined through the objective perspective of well-being defined by Stiglitz et al. [3].

Advanced analytics methods, such as artificial intelligence, machine learning and data-mining algorithms, are employed to analyse heterogeneous datasets and to support the identification, prediction and resolution of complex problems [4, 5, 6]. Previous research has demonstrated the significant potential of advanced analytics and artificial intelligence in various areas of clinical health care [7-11]. However, challenges in using analytics include data visualisation and practical applicability [8], lack of user understanding and ethical concerns [12, 13, 14], and technical limitations [8, 11, 15].

Co-creation is considered part of a broader concept of co-approaches, which utilise participatory and collaborative methods to leverage data in healthcare contexts. It brings together various stakeholders as active and equal partners, working with different types of data and employing creative methods to understand problems and develop solutions [16]. Co-creation is described as a process involving co-ideation, co-design, co-implementation and co-evaluation [17].

Although the possibilities of analytics are widely recognised in healthcare [7, 8, 18], its use in identifying regional wellbeing among families has been far less studied. The

---

<sup>1</sup> Corresponding Author: Marika Helkiö, [marika.helkio@sata.fi](mailto:marika.helkio@sata.fi).

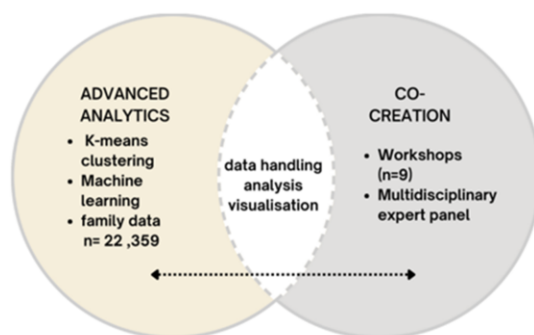
aim of this paper is to provide an overview of the analysis process of family register data and to describe the role of co-creation in this process.

## 2. Material and Methods

In Finland, basic services for families are organised through family centres within wellbeing services counties [19]. This analysis was conducted in one such county, where services for families are structured into four family centres. The study employed a complete study population sample comprising all individuals born between January 1, 2005, and December 31, 2023, as well as all other persons living in the same household, including siblings, cohabiting individuals, grandparents, parents, and other household members.

The Act on the Secondary Use of Social and Health Data regulates the conditions for using social and healthcare data for research purposes [20–21]. For this study, registry data were obtained from the Finnish Social and Health Data Permit Authority (Findata) under an approved data permit. The dataset included information from 11 national social and healthcare registers covered by the Act (Digital and Population Data Services Agency; Kela; Finnish Institute for Health and Welfare; and Finnish Centre for Pensions) as well as two registers outside the Act's scope (the Tax Administration and Statistics Finland), covering the years 2020–2023.

The data scientist analysed the dataset in Kapseli, which is a secure remote processing environment provided by Findata for handling individual-level data [20]. The variables included in the study were selected based on prior expert work [3, 22]. From the combined register datasets, 168 variables were extracted for each individual. The data scientist employed k-means clustering [23–25], in which families with similar wellbeing profiles are grouped into the same cluster (see Figure 1). Cluster analysis provides more detailed and granular anonymised information about the wellbeing of families. The clusters were further grouped into segments, offering a simplified and higher-level overview of family wellbeing.



**Figure 1.** Co-creation approaches to the analysis of family wellbeing data.

In accordance with the principles of co-creation [16], six workshops were organised during the data analysis phase, each attended by five participants. These participants had expertise in data processing, research, informatics, family service systems and leadership. The data scientist and one of the participants had prior experience with similar analyses [22]. In addition, three further workshops were organised, which—alongside the previously mentioned participants—included an expert in data visualisation. Each workshop was conducted remotely using the Microsoft Teams

application and lasted two hours. The purpose of the workshops was to support the success of the data scientist's analysis by incorporating diverse expertise, with the aim of ensuring the accuracy of the results and their practical applicability through consideration of end users' needs.

### **3. Results**

During the co-creation workshops, participants addressed matters related to data processing, analytical procedures and visualisation, while ensuring compliance with data security and privacy requirements. For example, postal code areas were aggregated to protect the privacy of families living in regions with very small populations. Although the analytical methods were familiar to the data scientist, the analysis required not only technical skills but also local contextual knowledge and, in particular, the operational experts' deep understanding of the study population. Through co-creation, the analysis incorporated the region's service structure, the selected analytical perspectives, and needs-based visualisation.

Using co-creation, 40 clusters and 8 segments were selected from families in the wellbeing services county, taking into account both the number of families in the region and its dispersed settlement structure. The data scientist proposed suitable numbers of clusters based on decisions made in the previous analysis [22]. An excessively high number of clusters produces fragmented and difficult-to-interpret information, whereas too few clusters exclude important details.

The integration and analysis of datasets from different registry sources resulted in a final study population of 22,359 families. Family wellbeing is examined across several wellbeing dimensions, including health, education, personal activities and work, voice in society, social connections and relationships, material living conditions, environment and sense of security [3]. Clusters and segments can also be viewed across multiple geographical levels: wellbeing services county (1), Family Centre area (4), municipality (16) and postal code areas (approximately 200). The output dataset includes four Excel-format files for each level, as mentioned above, as well as documentation for the dataset (guidance on the use of the files in the analysis and descriptions of the indicators and variables). The analysis also included two-dimensional visualisations, in which clusters and segments are shown on an XY-plane based on two selected wellbeing dimensions [3].

The end users of the wellbeing information produced in this analysis include professionals working with families in the region. To support these end users, a one-slide verbal and visual description of each cluster and segment was created in Microsoft PowerPoint, using radar charts and a bar chart with overlaid markers to illustrate the typical wellbeing patterns of the families within them. These cards can be used in practice in various ways, such as identifying differences in wellbeing, guiding the allocation of expertise or resources, or recognising specific wellbeing phenomena. In addition, a design plan was developed for implementing a Power BI solution, enabling interactive examination of the dataset from multiple perspectives. Because regional actors have different information needs, an interactive display is essential.

### **4. Discussion**

Advanced analytics can be used to identify meaningful family types within multidimensional register-based wellbeing data. The findings of this analysis support service planning, resource allocation and the identification of vulnerable groups. The

co-creation approach enriched the analysis by providing contextual understanding, addressing challenges, and generating new ideas (see also [16]). Similar features of co-creation were observed as those identified in [16], including shared goal-setting, equitable collaboration, shared decision-making, valuing diverse expertise, flexibility, a safe working environment, and long-term commitment. Co-creation approach thereby enabled the successful implementation of the analysis, while also ensuring that the needs of end users were systematically incorporated into the analytical process (see also [12, 14]).

Processing the data within the Kapseli environment proved to be a secure solution, enabling anonymised outputs to be generated in compliance with legal requirements [22]. Several challenges emerged during the analysis, including interpreting the AI-generated clusters, combining data into sufficiently large regional groups, ensuring anonymity, and addressing ethical considerations (see also [8, 13]), as well as the need for effective data visualisation (see also [9]). The paper provided further methodological insights and an improved understanding of visualising complex datasets. Previous research has shown that advanced analytics offers significant opportunities, although fully realising its potential still requires continued technological and methodological development [13-15].

Advanced analytics alone do not lead to better decisions or more effective services (see also [14]); its value emerges through co-creation with experts from different fields. The co-creation workshops and shared interpretation processes enable analytical results to be examined within real-world contexts, compared with everyday observations and translated into actionable insights for service planning and management. This participatory approach also supports ethically and socially sustainable knowledge use, as it ensures that analytical outputs remain connected to the lived realities of families and professionals. Co-creation thus serves as a bridge between analytics and practical decision-making—transforming complex data and analyses into a shared understanding and concrete development actions within family services.

## 5. Conclusions

Such an approach should become well established in the future development of social and health services, as it enables services to be targeted at those who need them most. An advantage in this analysis was the prior experience with similar work [22], as well as the diverse expertise brought by the co-creation process.

Further study will report findings on how wellbeing data on families are utilised in service planning within the Family Centres. In addition, forthcoming analyses will examine the substantive findings on regional wellbeing among families with children.

## Acknowledgements

We gratefully acknowledge the Ministry of Social Affairs and Health (STM) for supporting this research as part of Finland's Sustainable Growth Programme, funded through the EU's Next Generation EU recovery instrument. We also thank Findata for granting the data permit (THL/4632/14.02.00/2024) and enabling access to the data.

## References

- [1] OECD. *Measuring What Matters for Child Well-being and Policies*. Paris: OECD Publishing; 2021. doi:10.1787/e82fded1-en.
- [2] Amerijckx G, Humblet PC. Child well-being: what does it mean? *Children & Society*. 2014 Sep;28(5):1–13. doi:10.1111/chso.12003

- [3] Stiglitz JE, Sen A, Fitoussi JP. *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris: Commission on the Measurement of Economic Performance and Social Progress; 2009. p. 1–292
- [4] Elragal R, Elragal A, Habibipour A. Healthcare analytics: a literature review and proposed research agenda. 2023 Oct;1–15. doi:10.3389/fdata.2023.1277976.
- [5] Olowe K, Edoh NL, Zouo SJC, Olamijuwon J. Comprehensive review of advanced data analytics techniques for enhancing clinical research outcomes. *International Journal of Scholarly Research in Biology and Pharmacy*. 2024 Nov;5(1):8–17. doi:10.56781/ijrsbp.2024.5.1.0229.
- [6] Worku A, Arage FG, Kebede FB. Introduction to health care data analytics: an overview. *Discover Health Systems*. 2025 Dec 1;4(1). doi:10.1007/s44250-025-00291-x
- [7] Bradley PS. Implications of Big Data Analytics on Population Health Management. 2013 Jan;1(3):1–8. doi:10.1089/big.2013.0019.
- [8] Schulte T, Bohnet-Joschko S. How can Big Data Analytics support people-centred and integrated health services: a scoping review. *International Journal of Integrated Care*. 2022;22(2):1–18. doi:10.5334/ijic.5543.
- [9] Taipalus T, Isomöttönen V, Erkkilä H, Äyrämö S. Data analytics in healthcare: a tertiary study. *SN Computer Science*. 2023 Jan;4(1):1–14. doi:10.1007/s42979-022-01507-0.
- [10] Rani S, Jining D, Shah D, Xaba S, Singh PR. The potential application of artificial intelligence in healthcare and hospitals. *ITM Web of Conferences*. 2023 Jan;53:1–13. doi:10.1051/itmconf/20235301005.
- [11] Gombolay G, Gopalan N, Bernasconi A, Nabbout R, Megerian JT, Siegel B, Hallman-Cooper J, Bhalla S, Gombolay MC. Review of machine learning and artificial intelligence for the pediatric neurologist. *Pediatric Neurology*. 2023 Apr;141:42–51. doi:10.1016/j.pediatrneurol.2023.01.004.
- [12] Gillingham P. Can predictive algorithms assist decision-making in social work with children and families? *Child Abuse Review*. 2019 Mar 1 [cited 2026 Feb 7];28(2):114–26. doi:10.1002/car.2547
- [13] Shah N, Arshad A, Mazer MB, Carroll CL, Shein SL, Remy KE. The use of machine learning and artificial intelligence within pediatric critical care. *Pediatric Research*. 2023 Jan;93(2):405–12. doi:10.1038/s41390-022-02380-6.
- [14] Kotp MH, Ismail HA, Basyouny HAA, Aly MA, Hendy A, Nashwan AJ, et al. Empowering nurse leaders: readiness for AI integration and the perceived benefits of predictive analytics. *BMC Nursing*. 2025 Jan;24(1):1–13. doi:10.1186/s12912-024-02653-x.
- [15] Alanazi A. Clinicians' views on using artificial intelligence in healthcare: opportunities, challenges, and beyond. *Cureus*. 2023 Sep;15(9):1–11. doi:10.7759/cureus.45255.
- [16] Grindell C, Coates E, Croot L, O' Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. *BMC Health Services Research*. 2022 Dec;22(1):1–27. doi:10.1186/s12913-022-08079-y.
- [17] Pearce T, Maple M, Shakeshaft A, Wayland S, McKay K. What is the co-creation of new knowledge? A content analysis and proposed definition for health interventions. *International Journal of Environmental Research and Public Health*. 2025 Mar;17(7):1–20. doi:10.3390/ijerph17072229.
- [18] Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, Boccia S. Benefits and challenges of big data in healthcare: an overview of the European initiatives. *European Journal of Public Health*. 2019;29(3):23–7. doi:10.1093/eurpub/ckz168.
- [19] Finnish Institute for Health and Welfare. *Perhekeskus*. 2025. [cited 2026 Jan 22] Available from: <https://thl.fi/aiheet/lapset-nuoret-ja-perheet/sote-palvelut/perhekeskus>
- [20] Social and Health Data Permit Authority. *Kapseli*. 2025. [cited Jan 22] Available from: <https://findata.fi/en/kapseli/>
- [21] Ministry of Social Affairs and Health. *Act on the Secondary Use of Health and Social Data*. 2019. [cited Jan 22] Available from: <https://stm.fi/documents/1271139/1365571/...>
- [22] LEA. Lapsiperheiden edistynyt analytiikka (LEA) -hanke: Loppuraportti. IT-johtamisen toimintamalli. 2021. [cited Jan 22] Available from: [https://www.tampere.fi/sites/default/files/2022-05/Lapsiperheiden\\_edistynyt\\_analytiikka\\_loppuraportti.pdf](https://www.tampere.fi/sites/default/files/2022-05/Lapsiperheiden_edistynyt_analytiikka_loppuraportti.pdf).
- [23] Gao B, Yang Y, Gouk H, Hospedales TM. Deep clustering with concrete k-means. *CoRR*. 2019 Jan 1;abs/1910.08031:1–5 [cited 2026 Jan 25]. doi:10.48550/arXiv.1910.08031.
- [24] Sieranoja S. *Clustering with kNN graph and k-means*. Finland: University of Eastern Finland; 2020 [cited 2026 Jan 26]. Available from: <http://urn.fi/URN:ISBN:978-952-61-3645-5>
- [25] Szalkai B. An implementation of the relational k-means algorithm. *CoRR*. 2013 Jan 1;abs/1304.6899:1–8 [cited 2026 Jan 25]. doi:10.48550/arXiv.1304.6899.