International Journal for Quality in Health Care, 2021, 33(1), 1–8
doi: 10.1093/intqhc/mzab007
Advance Access Publication Date: 15 January 2021
Original Research Article

Original Research Article

Variability in the assessment of children's primary healthcare in 30 European countries

DANIELA LUZI[1], ILARIA ROCCO[2], OSCAR TAMBURIS[1], BARBARA CORSO[2], NADIA MINICUCI[2] and FABRIZIO PECORARO[1]

[1]National Research Council, Institute for Research on Population and Social Policies, via Palestro, 32, Rome, Lazio 00185, Italy, and [2]National Research Council, Neuroscience Institute, Via Giustiniani 2, Padua, Veneto 35128, Italy

Address reprint requests to: Fabrizio Pecoraro, National Research Council, Institute for Research on Population and Social Policies, via Palestro, 32, Rome, Lazio 00185, Italy. Tel: +39-06492724278; Fax: 39-06-4938-3724; E-mail: fabrizio.pecoraro@irpps.cnr.it

Received 20 April 2020; Editorial Decision 3 January 2021; Revised 12 December 2020; Accepted 13 January 2021

Abstract

Background: The high variability in the types and number of measures adopted to evaluate child-care across European countries makes it necessary to investigate country practices to identify trends in setting national priorities in the assessment of child well-being.
Objective: This paper intends to investigate country practices under the lens of variability to explore possible trends in setting national priority in the evaluation of childcare. In particular, it analyses variability considering to what extent this depends on the tendency of adopting a broad vision (i.e. selecting measures for a larger variety of aspects) or whether this is influenced by the choice of adopting an in-depth approach (i.e. using more measures to analyse a specific aspect)
Methods: An ad hoc questionnaire was administered to a national expert in each country and yielded 352 measures. To analyse variability, the breadth in the number of aspects considered was explored using a convergence index, while the depth in the distribution of measures in each aspect was investigated by computing a coefficient of variation. Countries were grouped by adopting a hierarchical clustering approach.
Results: There is a high variability across countries in the selection of measures that cover different aspects of childcare. Preferences in the distribution of measures are significant even at the domain level and in countries that use a limited number of measures and become more evident at the category and sub-category levels. The statistical analysis clusters countries in four main groups and two outliers. The in-depth distribution of measures focused on a specific aspect shows a homogeneous pattern, with the identification of two main groups of countries.
Conclusions: A limited set of measures are shared across countries hampering a robust comparison of paediatric models. The selection of measures shows that the evaluation is closely related to national priorities as resulting from the number and types of measures adopted. Moreover, a range of a reasonable number of measures can be hypothesized to address the quality of childcare under a multi-dimensional perspective.

Key words: measurement of quality, children, cross-country comparison, breadth, depth

Introduction

Since the publication of the World Health Report 2000 [1] and related criticism [2], country comparison has undergone important revisions in terms of measures to be used and procedures to gather reliable data with a renewed support towards the availability of robust data and their necessary harmonization. This applies in particular to the evaluation of child healthcare whose statistical invisibility largely hampers the identification of needs to be addressed to improve child well-being and well-becoming [3]. Efforts in the evaluation of child healthcare under a country comparison perspective are increasing [4, 5] along with studies measuring the strength of primary care [6], which even if not specifically focused on childcare outline important determinants of health.

Goals set internationally [7] and child-centric initiatives [8] contribute to underlining new areas of concern, areas that still need improvement and, not least, areas that have a major role in proposing metrics to monitor achievements [9, 10] under a country comparison perspective. Nevertheless, 'making children's lives visible' still remains one of the main priorities, which is also emphasized when reporting progress in countries that have signed European strategies [11, 12]. Moreover, the choice of appropriate metrics conveying a multidimensional appraisal of child healthcare is closely related to the selection of a *reasonable* number of measures [13, 14]. This is still an open debate that underscores the need for balancing between the use of an excessive and scarce number of measures to avoid performance paradox [13, 15] that may lead to weak correlation as well as to misinterpretation. Besides this, country comparison is crucial as it sheds light on priority setting and the breadth of aspects considered for the evaluation of quality of care. It also contributes to a cross-border learning from others.

This paper builds on the MOCHA (Models of Child Health Appraised) project that aims to describe and compare systems of primary care for children in 30 European countries adopting a multidisciplinary approach to analyse the different components that best address the needs of children and their families [16]. In this context, our contribution was to analyse common measures adopted to assess child healthcare in view of a country comparison.

Based on previous results [3, 14], this paper intends to further investigate country practices under the lens of variability to explore possible trends in setting national priority in the evaluation of childcare. It analyses variability considering whether this depends on the tendency of adopting a broad vision (i.e. selecting measures for a larger variety of aspects) or it is influenced by the choice of adopting an in-depth approach (i.e. using more measures to analyse a specific aspect). Appropriateness of measures used at country level is however out of scope in this study. This analysis may contribute to improving quality assurance of children's health services as well as reaffirming the necessity of protecting health and rights of this often-neglected important part of population [17].

Methods and materials

Data collection: the survey

A questionnaire was sent to MOCHA's appointed Country Agents (CAs), who have access to local documentation and rely on expert opinions, to gather information on measures used to evaluate child healthcare at national level. If an assessment agency was reported, CAs were asked to provide a list of all measures used to evaluate the quality of child healthcare along with any other official documents and/or links to websites of government bodies. This documentation was thoroughly examined, and measures used were extracted, harmonized and classified. To verify possible misinterpretation, the preliminary results of the analysis were sent back to the CAs who were asked to check and clarify potential ambiguities. Questionnaires and additional answers have been received up to May 2018. Among the 30 MOCHA countries, 27 have responded to the questionnaire with the exception of Luxembourg, Slovakia and France. Twenty-three countries were eligible for the analysis, as two countries (Greece and Malta) reported to have no agency for quality evaluation in primary care, and other two countries (Poland and Romania) perform accreditation procedures of primary and secondary service delivery, which cannot be aligned with measures adopted elsewhere.

The classification of results

As detailed in the work by Minicuci *et al.* [11], the whole set of measures collected were first classified in a two-level hierarchical map of domains, combining top-down and bottom-up approaches through an iterative analysis. The identification of domains had as a reference point the conceptual framework developed by Arah *et al.* [18] that has been variously adopted [12, 19–21] to perform cross-country comparison of health systems introducing a broader vision of health indicators that include societal and public health determinants. We built on this framework as we comprised context- and non-health-related measures, while we referred to the works by OECD, Rigby *et al.* and Bradshaw *et al.* [22–24] for a more centred child vision.

The resulting map (Figure 1) depicts the main areas of concern, that is the domains that group at high-level measures related to non-medical determinants of health (social, political, economic and environmental context—SPEEC; health-related behaviour—H-RB) and those more closely connected with health system performance indicators, traditionally based on the Donabedian framework (structure, process and outcome) [25]. The refinement of the five domains outlines 15 themes (hereafter categories) that are further detailed in 67 related aspects (sub-categories), which already show a broad spectrum of perspectives differently used to assess child healthcare.

The entire set of measures as well as their classification within the conceptual map is available in the Zenodo repository [26].

The measurement of the variability

The total number of 352 measures collected across the 23 responding countries are unequally distributed, from a minimum of 6 measures adopted in Iceland to a maximum of 130 in the UK. In our hypothesis, this variability may signal the different foci posed in the evaluation detecting national assessments more inclined to have a broad vision of childcare versus those that tend to analyse specific aspects in depth. Therefore, the frequency in the adoption of measures was explored within the domains, categories and sub-categories through the identification of the following:

- the breadth of child healthcare evaluation considering the number of categories and sub-categories covered by each country along with the number of countries that share similar features. This was obtained adopting a convergence index (see Appendix).
- the depth of childcare evaluation considering the distribution of measures in each category and sub-category along with the level of variability across countries. This was calculated through a coefficient of variation (see Appendix).

The results of these two interrelated analyses allowed us to explore, whether, within variation, countries could be clustered

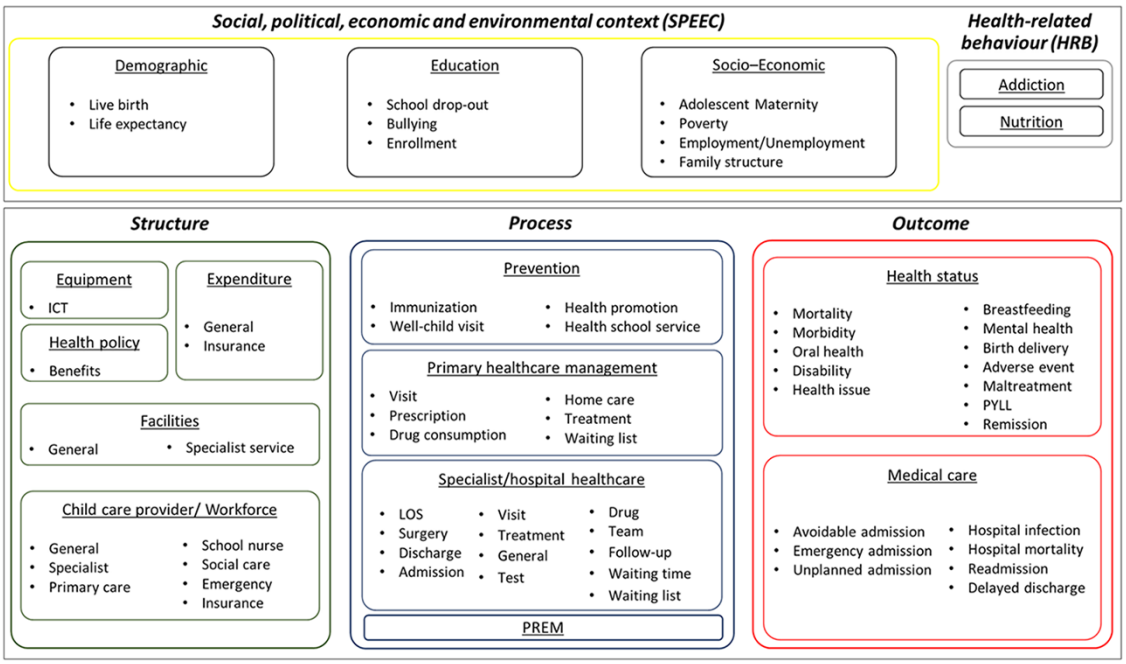

Figure 1 Conceptual map for the classification of measures [11].

according to similar behaviour in the selection of measures. This was achieved through a multivariant statistical analysis that adopts a hierarchical clustering approach and creates a complete data partition. In particular the Ward's method [27] based on the Euclidean distance/proximity matrix was applied. In this study, countries are clustered on the basis of the number of measures, categories and sub-categories and the related convergence index and coefficient of variation. The result is graphically represented using a dendrogram that depicts how countries are iteratively combined into groups on the basis of their dissimilarity (distance). A cut-off distance between groups (i.e. cut-tree) identifies the number and composition of clusters to be analysed.

Results

Analysis of the quantity and distribution of measures

As shown in Figure 2, the overall tendency is to privilege measures classified in the process (50.7%), outcome (33.0%) and structure (10.3%) domains. Country evaluation of non-medical determinants of health is rarely accomplished (3.1% and 2.9% for SPEEC and HR-B domains, respectively). This tendency could be explained considering that national health evaluation agencies generally rely on in a well-established administrative data flow generally aiming at cost control that traditionally refer to the Donabedian framework.

However, within this tendency, variations among countries are worth noting. Czech Republic is country where 67.7% ($n=6$) of measures are related to structure, followed by Croatia (33.3%, $n=6$), Bulgaria (27.3%, $n=6$) and Norway (26.3%, $n=5$). The wide adoption of measures related to process varies consistently from the Spanish highest proportion of measures (82.4%, $n=14$) to the lowest one in Bulgaria (13.6%, $n=3$). Iceland (33.3%, $n=2$) and Belgium (21.1%, $n=4$) are the countries with the highest percentage of measures related to SPEEC domain, while the Netherlands is the country with the highest percentage of H-RB measures (33.3%, $n=4$). Important to consider is that six countries (Austria, Germany, Finland, Ireland, Latvia and the UK) cover all the domains of the map suggesting efforts towards a multidimensional evaluation of childcare.

Considering common measures, the majority (87.1%, $n=306$) of them are reported by at most four countries, with no evident differences between domains. The lack of common measures is more evident, if we consider that almost half of the measures (44.6%, $n=157$) are reported by only one country. These results are in line with previous studies that consider the adoption of internationally suggested indicators [28, 29]. In our study, measures adopted by at least 9 countries are as follows: immunization rates/coverage (process domain, $n=13$ countries), infant mortality per 1.000 live births (outcome domain, $n=11$), number of live births (SPEEC domain, $n=10$), immunization coverage MMR (measles/mumps/rubella) (process domain, $n=9$), percentage of low-birth-weight newborns (outcome domain, $n=9$) and number of stillborn children per population (outcome domain, $n=9$). This confirms the worrying lack of common measures of children's healthcare quality in Europe.

Analysis of the breath of the thematic coverage

Based on previous studies [3, 11], Table 1 reports the number of categories, sub-categories and related measures adopted in each country along with the convergence index computed for these levels of the conceptual map. Results are presented by grouping countries in three main categories as shown by the arrows reported in the convergence index columns: (i) upward arrow denotes countries with a convergence in the adoption of core aspects shared with a relatively high number of countries (i.e. high value of the convergence index); (ii) rightward arrow denotes countries that adopt a mix of both core-shared and unique aspects (i.e. medium value of the convergence index) and (iii) downward arrow denotes countries that privileged the adoption of specific aspects not shared with other countries (i.e.

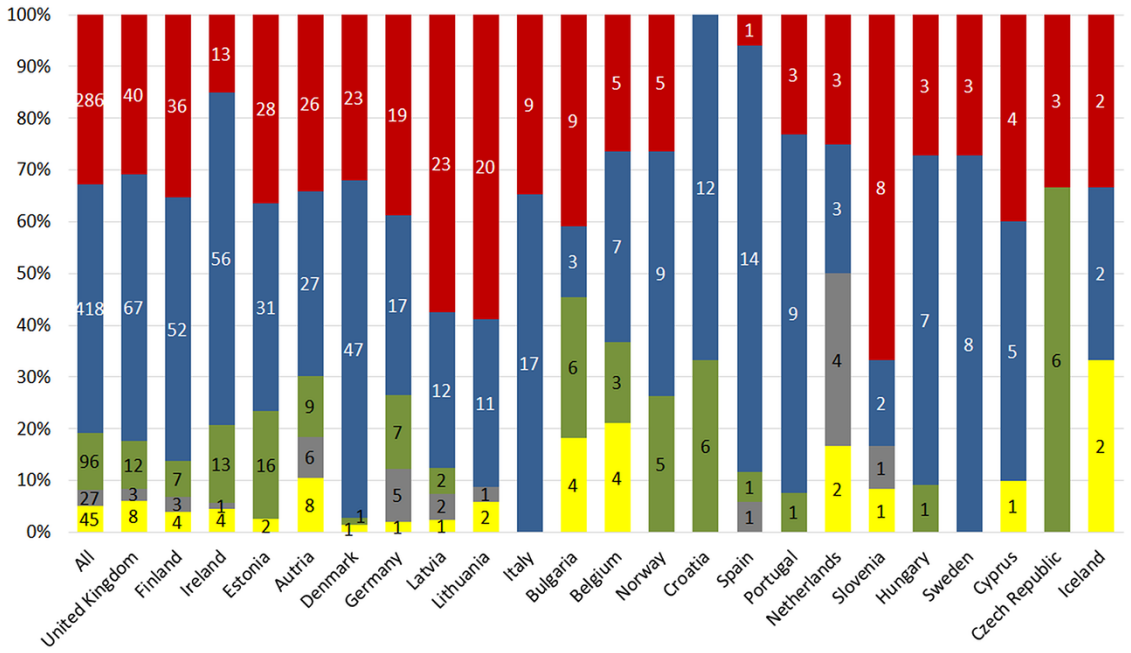

Figure 2 Distribution of the measures for each country, by domain; the number of measures adopted by each country in each domain is also reported.

Table 1 Number of categories, sub-categories and measures adopted by each country along with the relevant convergence indices

	Number of categories	CI categories	Number of sub-categories	CI sub-categories	Number of measures	CI measures
Iceland	4	↑ 17.8	4	↑ 13.0	6	↑ 5.2
Czech Republic	4	↓ 10.8	7	↓ 7.6	9	↓ 2.6
Cyprus	5	↑ 18.2	8	↑ 12.6	10	↑ 6.8
Hungary	5	↑ 17.2	9	→ 10.6	11	→ 3.4
Sweden	4	↑ 18.0	7	→ 10.6	11	↓ 2.7
Netherlands	7	↓ 13.1	10	→ 10.3	12	→ 5.0
Slovenia	5	→ 14.6	8	↑ 11.4	12	→ 3.7
Portugal	6	↑ 17.7	10	→ 11.1	13	→ 3.5
Spain	5	→ 15.4	8	→ 11.1	17	↓ 2.4
Croatia	6	→ 15.0	12	→ 9.5	18	→ 4.2
Belgium	9	→ 13.7	15	→ 10.7	19	↑ 5.2
Norway	6	↑ 16.8	15	→ 9.7	19	→ 4.6
Bulgaria	6	→ 14.2	11	→ 10.5	22	↑ 5.4
Italy	5	↑ 18.6	9	↑ 11.4	26	→ 4.2
Lithuania	8	→ 15.3	15	→ 10.7	34	→ 4.9
Latvia	9	→ 15.0	20	→ 9.9	40	→ 3.4
Germany	11	↓ 13.2	25	→ 9.5	49	→ 4.0
Denmark	7	↑ 16.4	25	↓ 8.2	72	↓ 1.4
Austria	13	↓ 12.9	34	↓ 8.1	76	→ 4.7
Estonia	11	↓ 12.9	34	↓ 8.2	77	→ 4.5
Ireland	13	↓ 12.1	36	↓ 7.9	87	→ 4.1
Finland	13	↓ 12.9	36	↓ 8.8	102	→ 4.5
U K	14	↓ 12.2	49	↓ 7.4	130	→ 4.2

Countries are reported in reverse order based on the number of measures.
CI, relevant convergence index

low value of the convergence index). To improve the readiness of the table, the results are ordered on the basis of number of measures reported by countries.

Figure 3 shows the dendrogram as a result of the clustering analysis highlighting how countries are grouped on the basis of the distance between clusters. The identified cut-tree (see cut-tree #1) distinguishes the group of countries according to the following main behaviours:

1. Countries that adopt a high number of measures (Austria, Estonia, Finland, the UK and Ireland) have low value of convergence indices covering a similar number of categories and sub-categories.

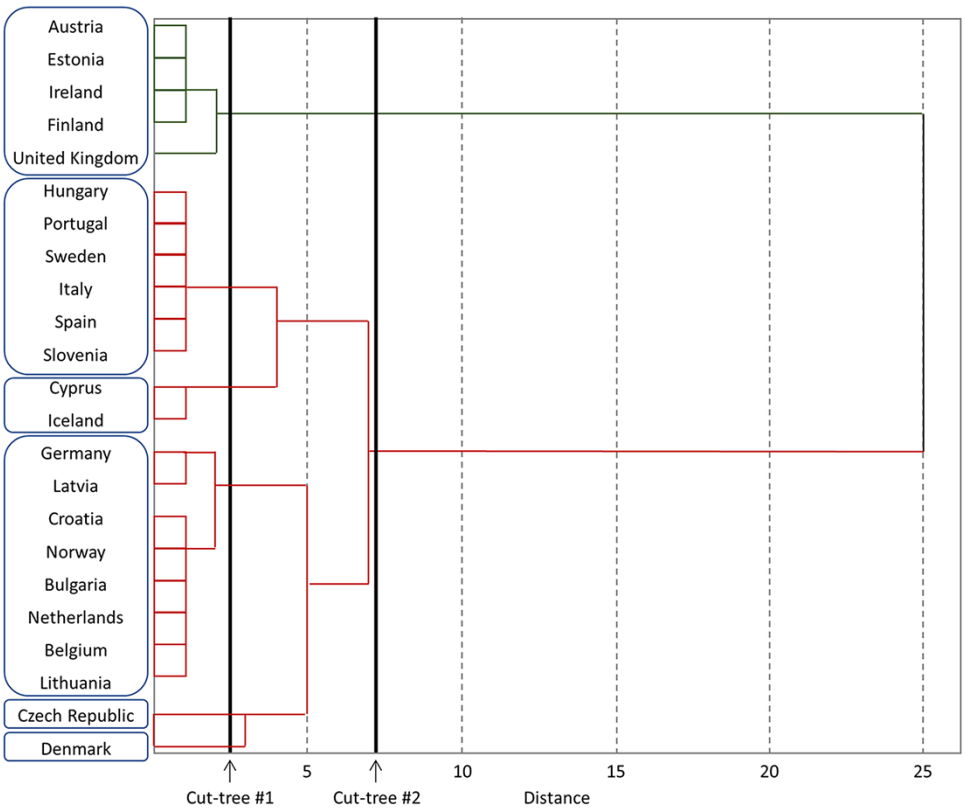

Figure 3 Dendrogram describing the hierarchical clustering analysis using the Ward's method based on the number of measures, categories and sub-categories as well as the relevant convergence indices; the solid black line (i.e. cut-tree) highlights the distance (dissimilarity) between clusters to specify which and how many clusters are finally identified.

Among them, the UK is the country that presents the highest number of measures covering the highest number of categories and sub-categories.

2. Countries (Cyprus and Iceland) with a limited number of measures that cover a low number of categories and sub-categories, but have relatively high values of the convergence indices.
3. Countries (Slovenia, Croatia, Germany, Lithuania and Latvia) that have similar values of convergence indices across the three levels, but have a different number of measures covering a varying number of categories and sub-categories.

In addition, two outliers are detected:

1. Denmark that adopts a high number of measures that are shared with few other countries and covers a limited number of common categories.
2. Czech Republic that has low values in the convergence indices in all levels of the map, despite a limited number of measures adopted.

Worth noting that at higher level of distance (see cut-tree #2), two main clusters may be detected. The first one is composed by Austria, Estonia, Finland, the UK and Ireland, which was the group more clearly identifiable adopting more than 70 measures to assess the quality of childcare. The second one represents all other countries with a high variation in the convergence indices using less than 50 measures.

Analysis of the depth

This part of the analysis explores whether each country homogeneously distributes measures within categories and sub-categories aiming at identifying tendencies towards the selection of a set of measures covering the same specific aspect of childcare from a different perspective (depth). Table 2 reports the coefficients of variation computed for each country in both the category and sub-category levels of the conceptual map. Results are ordered on the basis of the number of measures.

Generally, countries homogeneously distribute measures within the majority of categories and sub-categories. However, a clearly identifiable group of countries can be detected (Figure 4), which corresponds to the countries that have a broader vision in the evaluation of children's healthcare and use more than 70 measures (Austria, Denmark, Estonia, Finland, the UK and Ireland). The high number of measures adopted allows these countries to select also particular aspects and evaluate them under different perspectives using a different set of measures. For instance, Finland distributes its 102 measures in 36 sub-categories privileging a homogeneous distribution (from 1 to 4 measures) in 30 sub-categories but also detailing others with a consistent number of measures: admission ($n = 10$), birth delivery ($n = 7$), immunization ($n = 8$), morbidity ($n = 8$), mortality ($n = 10$) and well-child visit ($n = 6$). Similarly, Ireland focuses the attention on five sub-categories: admission ($n = 9$), immunization ($n = 7$), mortality ($n = 7$), waiting list ($n = 6$) and well-child visit ($n = 6$). The UK has the same pattern, but among the 49 sub-categories the focus is on the following: admission ($n = 13$), immunization ($n = 7$) and

Table 2 Number of categories and sub-categories adopted by each country along with the relevant coefficient of variation

	Number of categories	CV categories	Number of sub-categories	CV sub-categories	Number of measures
Iceland	4	↓0.4	4	↓0.4	6
Czech Republic	4	↓0.4	7	↓0.4	9
Cyprus	5	↓0.7	8	↓0.6	10
Hungary	5	↓0.7	9	↓0.4	11
Sweden	4	↓0.6	7	↓0.3	11
Netherlands	7	↓0.4	10	↓0.4	12
Slovenia	5	→1.1	8	→0.7	12
Portugal	6	→1.1	10	→0.7	13
Spain	5	→1.1	8	→0.6	17
Croatia	6	↓0.8	12	→0.6	18
Belgium	9	↓0.7	15	↓0.5	19
Norway	6	↓0.5	15	↓0.5	19
Bulgaria	6	↓0.7	11	↑0.9	22
Italy	5	↓0.5	9	→0.7	26
Lithuania	8	→1.1	15	↑1.0	34
Latvia	9	↑1.5	20	→0.7	40
Germany	11	→1.1	25	→0.7	49
Denmark	7	→1.1	25	↑1.1	72
Austria	13	↑1.2	34	↑0.9	76
Estonia	11	→0.9	34	→0.8	77
Ireland	13	↑1.3	36	↑0.9	87
Finland	13	↑1.2	36	↑1.0	102
U K	14	↑1.3	49	↑1.0	130

CV, relevant coefficient of variation

mortality ($n = 14$). Similar to this group of countries, Denmark distributes a high number of measures on a relatively limited number of aspects, privileging in particular the evaluation related to specialist visit ($n = 15$) selecting measures to monitor asthma, diabetes and attention deficit hyperactivity disorder (ADHD). Note that Denmark is the only country that select measures related to ADHD together with other mental health aspects such as depression which is evaluated considering the consuption of antidepressants among children and adolescents.

Discussion and conclusions

This paper posed its major focus on the issue of variability in the use of number and types of measures adopted to evaluate childcare across the MOCHA countries. This analysis considered in particular two interrelated points of view that may determine countries' choice: a breadth and a depth vision in the evaluation of childcare. The first one tends to spread measures within a wide spectrum of different aspects of child well-being, while the second one is more inclined to distribute a set of measures to evaluate a specific area of concern. The identification of national practices towards one of these tendencies, considering the different combinations of them, can contribute to interpreting this high variability. The adoption of a convergence index and a coefficient of variation applied to categories, sub-categories and measures helped us outlining important differences across countries. Moreover, the analysis of correlation that prompted the hierarchical clustering supported the identification of common tendencies, albeit the differences in the number of measures adopted by each MOCHA country.

One of the first results of this analysis confirms that a very limited set of measures are shared across countries hampering a robust comparison of paediatric models of childcare. This is shown already at the baseline level, if we consider that the majority of measures are shared at most by four countries and that among the six measures adopted by at least nine countries, the most adopted one (i.e. immunization rate) is reported by 13 countries. Preferences in the distribution of measures are significant even at the domain level and even in countries that use a limited number of measures. These differences become more evident at the category level and more remarkably at the sub-category one. The convergence index is relatively higher only in some of the countries that adopt a limited number of measures (6–10 measures) ranging at the sub-category level from 12.6 to 11.1, while similar values in both the convergence index and the number of measures produce in other countries the opposite results, indicating a low number of measures shared. Similarly, countries with a low-to-medium number of measures do not always present low values in the convergence index, and variability generally increases in countries with medium number of measures used. Noteworthy is the clear identification of countries, more evident and confirmed in the hierarchical clustering analysis, that adopt a consistent number of measures and have a low convergence index (Austria, Estonia, Finland and to a certain extent the UK). Moreover, outlier countries, such as Denmark and the Czech Republic, outline two opposite examples at the extreme of this wide spectrum of variability.

The analysis of national practices in an in-depth distribution of measures shows a relatively more homogeneous pattern composed of two groups of countries. A consistent group comprises countries that use both a limited (6–19) and a medium number (20–49) of measures. The second group comprises countries that more widely distribute their measures across categories and subcategories (Austria, Estonia, Finland and the UK) with the addition of Denmark. Note the in-depth vision of these countries generally refers to aspects that can be considered core in health system evaluation, such as immunization, mortality and hospital admission [13], which certainly have a long-standing tradition and a consolidated, but also country-specific data acquisition. The addition of Denmark in this group is worth noting

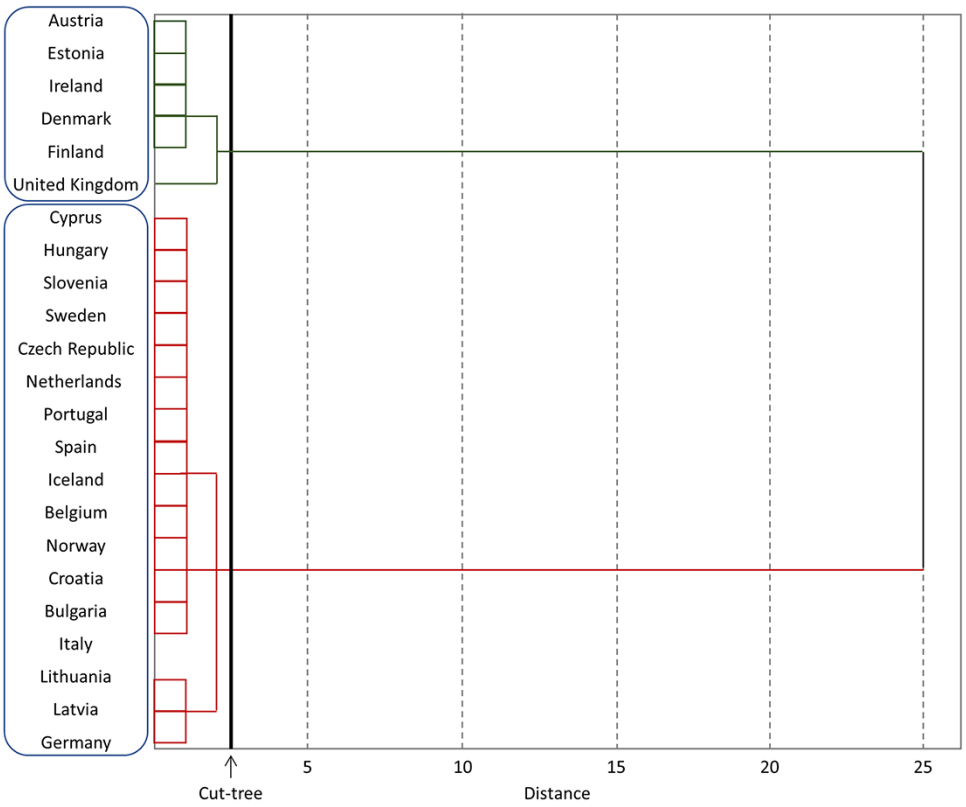

Figure 4 Dendrogram describing the hierarchical clustering analysis using the Ward's method based on the number of measures, categories and sub-categories as well as the relevant coefficients of variation; the solid black line (i.e. cut-tree) highlights the distance (dissimilarity) between clusters to specify which and how many clusters are finally identified.

as this country privileges aspects not considered elsewhere—in some cases emerging issues as suggested by European strategy, such as mental health. The thresholds identified (70 measures) in the analysis of both breadth and depth may lead us to suggest a range of reasonable number of measures suitable to target child healthcare under a multidimensional perspective. Based on this cut-off, a grey zone can be identified, which delimits countries with a medium number of measures (less than 50) and those with a consistent number of measures (more than 70).

Even if further analyses are needed, in our hypothesis, the reasonable number of measures could fall within this range. This could also give space to include emerging aspects and/or select country-specific concerns.

Acknowledgements

We thank Prof. Michael Rigby (Emeritus Professor of Health Information Strategy, Keele University, UK) for his valuable and precious comments and suggestions.

Data availability statement

The data underlying this article are available in Zenodo at https://www.doi.org/10.5281/zenodo.3339345.

Funding

This work was supported by the European Commission through the Horizon 2020 Framework (grant agreement number 634201). The sole responsibility for the content of this project lies with the authors. It does not necessarily reflect the opinion of the EU. The European Commission is not responsible for any use that may be made of the information contained therein.

References

1. World Health Organization. The World Health Report 2000: health systems: improving performance. World Health Organization, 2000.
2. McKee M. The World Health Report 2000: 10 years on. *Health Policy Plan* 2010;**25**:346–8.
3. Rocco I, Corso B, Luzi D *et al.* The conundrum of measuring children's primary health care. In: Blair M, Rigby M, Alexander D (eds). *Issues and Opportunities in Primary Health Care for Children in Europe*. United Kingdom: Emerald Publishing Limited, 2019,159–78.
4. van Esso D, Del Torso S, Hadjipanayis A *et al.* Primary-Secondary Working Group (PSWG) of European Academy of Paediatrics (EAP), paediatric primary care in Europe: variation between countries. *Arch Dis Child* 2010;**95**:791–5.
5. Wolfe I, Thompson M, Gill P *et al.* Health services for children in western Europe. *Lancet* 2013;**381**:1224–34.
6. Kringos DS, Boerma WGW, Hutchinson A *et al. Building Primary Care in a Changing Europe*. Copenhagen, Denmark: Observatory Studies Series, 2015,38.
7. *Transforming our World: The 2030 Agenda for Sustainable Development*. New York, NY, USA. A/RES/70/1 2015 http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E (6 April 2020, date last accessed).
8. WHO Regional Office for Europe. *Investing in Children: The European Child and Adolescent Health Strategy 2015–2020 (EUR/RC64/12)*. Copenhagen: World Health Organization, 2014.

9. Aleman-Diaz AY, Backhaus S, Siebers LL *et al.* Child and adolescent health in Europe: monitoring implementation of policies and provision of services. *Lancet Child Adolesc Health* 2018;**2**:891–904.
10. World Health Organization. *Situation of Child and Adolescent Health in Europe*. Copenhagen: WHO, 2018. ISBN 978-92-890-5348-8
11. Minicuci N, Corso B, Rocco I *et al.* Innovative measures of outcome and quality of care in child primary care models. 2017. http://www.childhealthservicemodels.eu/wp-content/uploads/2015/09/D7-Identification-and-Application-of-Innovative-Measures-of-Quality-and-Outcome-of-Models.pdf (6 April 2020, date last accessed).
12. Kringos D, Nuti S, Anastasy C *et al.* Re-thinking performance assessment for primary care: opinion of the expert panel on effective ways of investing in health. *Eur J Gen Pract* 2019;**25**:1–8.
13. Nuti S, Vainieri M, Vola F. Priorities and targets: supporting target setting in healthcare. *Public Money Manag* 2017;**37**:277–84.
14. Van Thiel S, Leeuw FL. The performance paradox in the public sector. *Public Perform Manag* 2002;**25**:267–81.
15. Blair M, Rigby M, Alexander D. *Issues and Opportunities in Primary Health Care for Children in Europe*. United Kingdom: Emerald Publishing Limited, 2019.
16. Rigby M, Deshpande S, Luzi D *et al.* The invisibility of children in data systems. In: Blair M, Rigby M, Alexander D (eds). *Issues and Opportunities in Primary Health Care for Children in Europe*. The Netherlands: Emerald Publishing Limited, 2019,129–58.
17. Rigby MJ, Deshpande S, Blair M. Child health research and planning in Europe disadvantaged by major gaps and disparities in published statistics. *Eur J Public Health* 2020;**30**:693–697.
18. Arah OA, Westert GP, Hurst J *et al.* A conceptual framework for the OECD health care quality indicators project. *Int J Qual Health Care* 2006;**18**:5–13.
19. Kringos DS. *The Strength of Primary Care in Europe*. Dissertation. The Netherlands: University Utrecht, 2012.
20. Mattke S, Epstein AM, Leatherman S. The OECD health care quality indicators project: history and background. *Int J Qual Health Care* 2006;**18**:1–4.
21. AHRQ (Agency for Healthcare Research and Quality). "Doing better for children" a monograph. OECD (http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/chtoolbx/why/index.html 6 April 2020, date last accessed).
22. OECD. *Doing Better for Children*. 2009. http://www.oecd.org/els/social/childwellbeing 6 April 2020, date last accessed.
23. Rigby MJ, Köhler LI, Blair ME *et al.* Child health indicators for Europe: a priority for a caring society. *Eur J Public Health* 2013;**13**:38–46.
24. Bradshaw J, Petra H, Dominic R *Comparing Child Well-Being in OECD Countries: Concepts and Methods*. 2007. (https://www.unicef-irc.org/publications/464-comparing-child-well-being-in-oecd-countries-concepts-and-methods.html 6 April 2020, date last accessed).
25. Donabedian A. The quality of care: how can it be assessed? *JAMA* 1988;**2**:1743–8.
26. Luzi D, Rocco I, Tamburis O *et al.* Corresponding spreadsheet to the paper 'variability in the assessment of childcare in 30 European countries'. 2020.
27. Ward JH. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 1963;**58**:236–44.
28. Rotar AM, Van den Berg MJ, Kringos DS *et al.* Reporting and use of the OECD Health Care Quality Indicators at national and regional level in 15 countries. *Int J Qual Health Care* 2016;**28**:398–404.
29. Braithwaite J, Hibbert P, Blakely B *et al.* Health system frameworks and performance indicators in eight countries: a comparative international analysis. *SAGE Open Med* 2017;**5**:2050312116686516.

Appendix

The breadth of childcare evaluation is assessed considering the convergence index calculated as follows:

$$CI = \frac{\sum_{i=1}^{k} n_i}{k}$$

where:

- k is the number of measures/categories/sub-categories adopted by the relevant country;
- n_i is the number of MOCHA countries reporting the measure/covering the category/sub-category i.

The depth of childcare evaluation is assessed considering the coefficient of variation calculated on the basis of the following equation:

$$CV = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\mu}\right)^2}$$

where:

- n is the number of categories/sub-categories covered by the relevant country;
- x_i is the number of measures adopted to describe the i-th categories/sub-categories;
- μ is the average number of measures adopted to describe the n categories/sub-categories covered by the relevant country.