Magnitude of terminological bias in international health services research: a disambiguation analysis in mental health

Author for correspondence:
M. R. Gutierrez-Colosia,
E-mail: menciaruiz@uloyola.es

M. R. Gutierrez-Colosia[1,2] ![ORCID], P. Hinck[3], J. Simon[4,5], A. Konnopka[3,6], C. Fischer[4], S. Mayer[4], V. Brodszky[7], L. Hakkart-van Roijen[8], S. Evers[9,10], A. Park[11], H. H König[3], W. Hollingworth[12], J. A Salinas-Perez[13,14], the PECUNIA Group[4] and L. Salvador-Carulla[14,15]

[1]Department of Psychology, Universidad Loyola Andalucía, Seville, Spain; [2]Scientific Association PSICOST, Seville, Spain; [3]Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [4]Department of Health Economics, Center for Public Health, Medical University of Vienna, Vienna, Austria; [5]Department of Psychiatry, University of Oxford, Oxford, UK; [6]Department of Psychology, Medical School Hamburg, Hamburg, Germany; [7]Department of Health Economics, Corvinus University of Budapest, Budapest, Hungary; [8]Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Burgemeester Oudlaan 50, PO Box 1738, 3000 DR, Rotterdam, The Netherlands; [9]Department of Health Services Research, Care and Public Health Research Institute (CAPHRI), Faculty of Health, Medicine and Life Sciences (FHML), Maastricht University, Maastricht, The Netherlands; [10]Trimbos, Netherlands Institute of Mental Health and Addiction, Da Costakade 45, 3521 VS, Utrecht, The Netherlands; [11]Department of Health Policy, Care Policy and Evaluation Centre, London School of Economics and Political Science, London, UK; [12]Health Economics Bristol, Population Health Sciences, University of Bristol, Bristol, UK; [13]Department of Quantitative Methods, Universidad Loyola Andalucía, Sevilla, Spain; [14]Faculty of Health, Health Research Institute, University of Canberra, Canberra, Australia and [15]National Centre for Epidemiology and Population Health, Faculty of Health and Medicine, Australian National University, Canberra, Australia

Abstract

Aims. Health services research (HSR) is affected by a widespread problem related to service terminology including non-commensurability (using different units of analysis for comparisons) and terminological unclarity due to ambiguity and vagueness of terms. The aim of this study was to identify the magnitude of the terminological bias in health and social services research and health economics by applying an international classification system.

Methods. This study, that was part of the PECUNIA project, followed an ontoterminology approach (disambiguation of technical and scientific terms using a taxonomy and a glossary of terms). A listing of 56 types of health and social services relevant for mental health was compiled from a systematic review of the literature and feedback provided by 29 experts in six European countries. The disambiguation of terms was performed using an ontology-based classification of services (Description and Evaluation of Services and DirectoriEs – DESDE), and its glossary of terms. The analysis focused on the commensurability and the clarity of definitions according to the reference classification system. Interrater reliability was analysed using κ.

Results. The disambiguation revealed that only 13 terms (23%) of the 56 services selected were accurate. Six terms (11%) were confusing as they did not correspond to services as defined in the reference classification system (non-commensurability bias), 27 (48%) did not include a clear definition of the target population for which the service was intended, and the definition of types of services was unclear in 59% of the terms: 15 were ambiguous and 11 vague. The κ analyses were significant for agreements in unit of analysis and assignment of DESDE codes and very high in definition of target population.

Conclusions. Service terminology is a source of systematic bias in health service research, and certainly in mental healthcare. The magnitude of the problem is substantial. This finding has major implications for the international comparability of resource use in health economics, quality and equality research. The approach presented in this paper contributes to minimise differentiation between services by taking into account key features such as target population, care setting, main activities and type and number of professionals among others. This approach also contributes to support financial incentives for effective health promotion and disease prevention. A detailed analysis of services in terms of cost measurement for economic evaluations reveals the necessity and usefulness of defining services using a coding system and taxonomical criteria rather than by 'text-based descriptions'.

![CAMBRIDGE UNIVERSITY PRESS logo]

Introduction

Health services research (HSR), health economics and financing, and research of quality and equality in healthcare require comparable data on service provision (Husereau *et al.*, 2013; Raine *et al.*, 2016). However, the reporting methods can differ substantially, and HSR faces significant problems regarding the terminology of services (Salvador-Carulla *et al.*, 2013), its implication for the measurement of resource utilisation (Thorn *et al.*, 2013) and its monetary valuation (Moreno *et al.*, 2008; Barnett, 2009). Terminology is defined as 'a set of designations belonging to one special language'(Roche, 2012), and its main purpose is to eliminate ambiguity from technical languages by means of standardisation. The main terminological problems in scientific research are unclarity due to ambiguity or vagueness of scientific terms. Terminological ambiguity exists when a term (the dyad of a name and its definition) can reasonably be interpreted in more than one way (e.g. two different codes of a reference classification system can be assigned to the same entity), and vagueness occurs when a word or phrase is underspecified and therefore admits borderline cases or relative interpretation (e.g. typically more than three codes can be assigned to the defined entity) (Castelpietra *et al.*, 2021). Disambiguation is the act of making something clear and this takes place in science by using taxonomies, for example, an ontology-based classification coding system and its related glossary of terms or dictionary. Ontoterminology is the discipline that studies disambiguation of technical and scientific terms using classifications, glossaries of terms and the related standard instruments (Castelpietra *et al.*, 2021).

The two major terminological problems in HSR are the non-commensurability bias and terminological unclarity. Non-commensurability is due to research involving different units of analysis that are not comparable like-with-like. For example, it occurs when the costs of an outpatient psychotherapy unit (i.e. a 'service') are compared to the costs of psychotherapy as an 'intervention' in another setting. The problem of terminological unclarity is also widespread. For example, the term 'service' can refer to a range of elements such as the provider, the facility, an organisational unit within the facility or a combination of functions, programmes and resources provided in this clinical unit (Salvador-Carulla *et al.*, 2013). Another problem refers to the lack of a formal definition of the target population the service has been designed for (i.e. diagnosis group and age group) (Salinas-Pérez *et al.*, 2020), and the variability in the typology of services depending on location (different areas, regions, countries, etc.) and time of evaluation (Salvador-Carulla *et al.*, 2015).

Twenty years ago, Maciejewski *et al.* identified significant terminology problems in the methods for HSR. To overcome these problems, these authors produced a list of terms commonly used in HSR methods following a scoping review of the literature and internal and external expert consultations (Maciejewski *et al.*, 2002). These terminology problems have been also described in mental health service evaluation (Salvador-Carulla *et al.*, 1999; Salvador-Carulla and Hernández-Peña, 2011) and are critical in the standardisation of international resource use measurement (RUM) instruments (Thorn *et al.*, 2013; Noben *et al.*, 2016).

Despite previous efforts, the terminology problems in HSR remain largely unnoticed and unaccounted for. For instance, in the USA, the Institute of Medicine (IoM) prioritised different areas of comparative effectiveness research but did not mention this source of systematic bias (Iglehart, 2009). Likewise, there is a substantial degree of variation in the applied valuation methods in health economic studies and guidelines in Europe, but the terminological variability is rarely being mentioned (van Lier *et al.*, 2018; Mayer *et al.*, 2020). The World Health Organization's Family of International Classification (WHO-FIC) has recently incorporated the International Classification of Health Interventions (ICHI) (Fortune *et al.*, 2018) to its list. However, the classification of the services where these interventions occur is still missing, and alternative solutions such as the Service Availability and Readiness Assessment (SARA) (O'Neill *et al.*, 2013) are too vague and broad to be used in comparative effectiveness or for disambiguation. The System of Health Accounts (SHA 2.0) (OECD Eurostat WHO, 2017) includes separate components of Health Providers and Health Functions but it lacks a reference taxonomy and a standard glossary and shows major consistency problems (Salvador-Carulla and Hernández-Peña, 2011).

The lack of specific studies on terminological in HSR is surprising. It could be attributed to the complexity of the analysis required to measure the extent of this problem (Maciejewski *et al.*, 2002); and the absence of a proper framework of reference until very recently. In the last decade, ontoterminology has been proposed in information technologies (Roche, 2012) and adapted to disambiguation in HSR (Castelpietra *et al.*, 2021). Apart from providing an adequate framework for the analysis of terms in a given field, a classification using a hierarchical taxonomy with a coding system provides a reference framework to code definitions as acceptable, or as ambiguous, vague or confusing (i.e. wrong or mistaken) in a reproducible way (Castelpietra *et al.*, 2021).

The aim of the study was to identify the magnitude of the bias of non-commensurability and terminological unclarity in health and social services research by applying an international classification system for coding human services and adapting it as needed to the newly emerging requirements for health economics research from a societal perspective. A complementary objective of this study was to demonstrate the usability of the ontoterminology approach to disambiguation in complex topics in healthcare research.

Methods

This study was part of the PECUNIA project (ProgrammE in Costing, resource use measurement and outcome valuation for Use in multi-sectoral National and International health economic evaluAtions) conducted from 2018 to 2021. PECUNIA was aimed at developing standardised multi-sectoral, multi-national and multi-person RUM instruments, unit cost valuation templates, reference unit costs and outcome assessment tools to improve the methodology and comparability of economic evaluations in the European Union with a special focus on mental health (Mayer *et al.*, 2022). The PECUNIA consortium coordinated by the Medical University of Vienna consisted of ten partners for health economics and health systems research, located in six European countries (Austria, Germany, Hungary, Spain, The Netherlands and UK) (The PECUNIA group, 2018). Due to their high disease burden and economic relevance, three mental disorders (depression, schizophrenia and posttraumatic stress disorder) were chosen as reference disorders to analyse the applicability of the newly developed methods and tools. This study concentrated on the disambiguation of services in the health

and social care cluster relevant for mental health and was carried out in parallel to other activities of the project.

Procedure

This ontoterminology study was performed using the Standards for Quality Improvement Reporting Excellence (SQUIRE 2.0) (Ogrinc *et al.*, 2016). It used a mixed-methods approach and followed a multistep process to assess the clarity of terms on health and social services. This process included three steps: (I) a systematic review to produce a preliminary listing of service terms based on scientific and grey literature; (II) an expert survey and a consecutive revision for the production of the final listing; (III) disambiguation of the terms included in the final listing adapting the previously tested method of Maciejewski *et al.* (2002). Steps I and II are fully described elsewhere (Fischer *et al.*, 2022; Mayer *et al.*, 2022).They involved three working groups and two expert panels. Working group A included three experts in health economics from the University of Hamburg (PH, AK, CD) that led step I and participated in step II. Working group B comprised three members from the Medical University of Vienna (JS, CF, SM) who led step II and participated in step I. Internal expert panel included the PECUNIA country leads and participated in steps I and II. Finally, the external expert panel was composed of 29 health and social service researchers, health economists, and planners from public agencies and other stakeholders in every participating country that provided an external validation of terms identified previously in step I.

Working group C consisted of two experts in health system terminology and coding from Psicost in Spain and Australia (MGC and LSC) who carried out the disambiguation analysis (step III). The whole process and the activities performed by the working groups and the expert panels are shown in Fig. 1.

Ontoterminology tools

We used an updated glossary of terms based on the Psicost and REFINEMENT glossaries (Montagni *et al.*, 2018) and an international classification of human services, the Description and Evaluation of Services and Directories (DESDE system) (Salvador-Carulla *et al.*, 2013) for the disambiguation of terms. The REFINEMENT glossary provides consensus-based operational definitions of the basic terms relevant for the disambiguation process in health services. DESDE has been used for the comparison of mental health service typologies across countries (Alonso-Solís *et al.*, 2020), analysis of disambiguation of complex terms in health care, such as psychotherapy, (Castelpietra *et al.*, 2021) and for the content analysis of the national classifications system compared to an international standard (Rosen *et al.*, 2020). Previous research has shown that the DESDE instrument scores high in feasibility, consistency, inter-rater reliability as well as face, content and construct validity (Salvador-Carulla *et al.*, 2013), as well as its applicability in health economics studies (Romero-Lopez-Alberca *et al.*, 2019).

These two related tools facilitate different types of disambiguation. On the one hand, the glossary of terms provides an operational definition of services for the identification of these units of analysis (commensurability) (Montagni *et al.*, 2018).

On the other hand, the DESDE uses its multiaxial coding system for clarity in the definition of terms. Two DESDE axes were used in the disambiguation process. The 'target' axis includes a code sub-thread to define the specific target population for whom the service is intended (age, gender, ICD coding, functioning and severity). The 'service' axis includes the specific code of the service type and its qualifiers. This code is based on the principal function of the service described as 'Main Types of Care' (MTCs). There are six main branches that describe the type of care: Residential care, Day care, Outpatient care, Self-help support, Information and Assessment, and Accessibility of care (Romero-Lopez-Alberca *et al.*, 2019). The DESDE hierarchical taxonomy includes 106 codes in five levels of granularity (main branch of care, acute/non-acute care, mobile/non-mobile care, physician/non-physician cover, intensity of care) labelled with an alphanumeric code. In this study, we requested for disambiguation only the two first digits of the label (letter + one number, e.g. O5 in Fig. 1), instead of using the full five levels of granularity of the MTC taxonomy.

In the example presented in Fig. 2, the code thread refers to the BSIC type called 'Assertive Community Treatment team' in the Basque Country (Spain) (García-Alonso *et al.*, 2019).

Disambiguation

Steps I and II yielded a list of 56 key services relevant to mental health based on extensive literature review and selected by international expert panels. Terms and definitions were classified into the following categories: accurate (the term could be classified using one code); ambiguous (the term was labelled with more than one – typically two – code), vague (the term could be coded with a series of codes), confusing (the term was wrong or incomplete according to the reference classification system as it required additional significant interpretation from the experts). Definitions were analysed at different levels. In level 1, the two raters confirmed that the definitions corresponded to services and not to another unit of analysis in HSR such as procedures, interventions or professionals, to ensure the commensurability of the terms included in the listing (Salvador-Carulla *et al.*, 2015). In level 2, the two raters analysed the information on the target population for which the service was intended. This included age, diagnosis group, functioning or other characteristics influencing health status and contact with health services (e.g. homelessness, domestic violence). Finally, level 3 of disambiguation included the definition of the service typology using DESDE taxonomy based on MTC codes. For further clarification, a full DESDE code was provided for every service. When problems in the definition of the service did not allow the assignment of a code, a prototype code was generated based on the interpretation made by the two evaluators. For example, in the list provided, 'outpatient healthcare service' is defined as a contact with the provider. This was categorised as confusing because 'contact' refers to an activity conducted by a professional. Codification was based on the most exemplary instance of that type of service according to expert analysis and interpretation. It has been underlined and written in italics (online Supplementary Table 1). Interobserver agreement was analysed using the standard interrater reliability analysis of Cohen's κ for categorical variables; results were interpreted following Landis and Koch's criteria (1977). Data were analysed by using SPSS Statistics for Windows.

Ethics

The listing of terms and definitions used in this study did not require ethical approval or consents in the participating countries as data were obtained from a review of the scientific and grey literature and did not include information on individual patients.
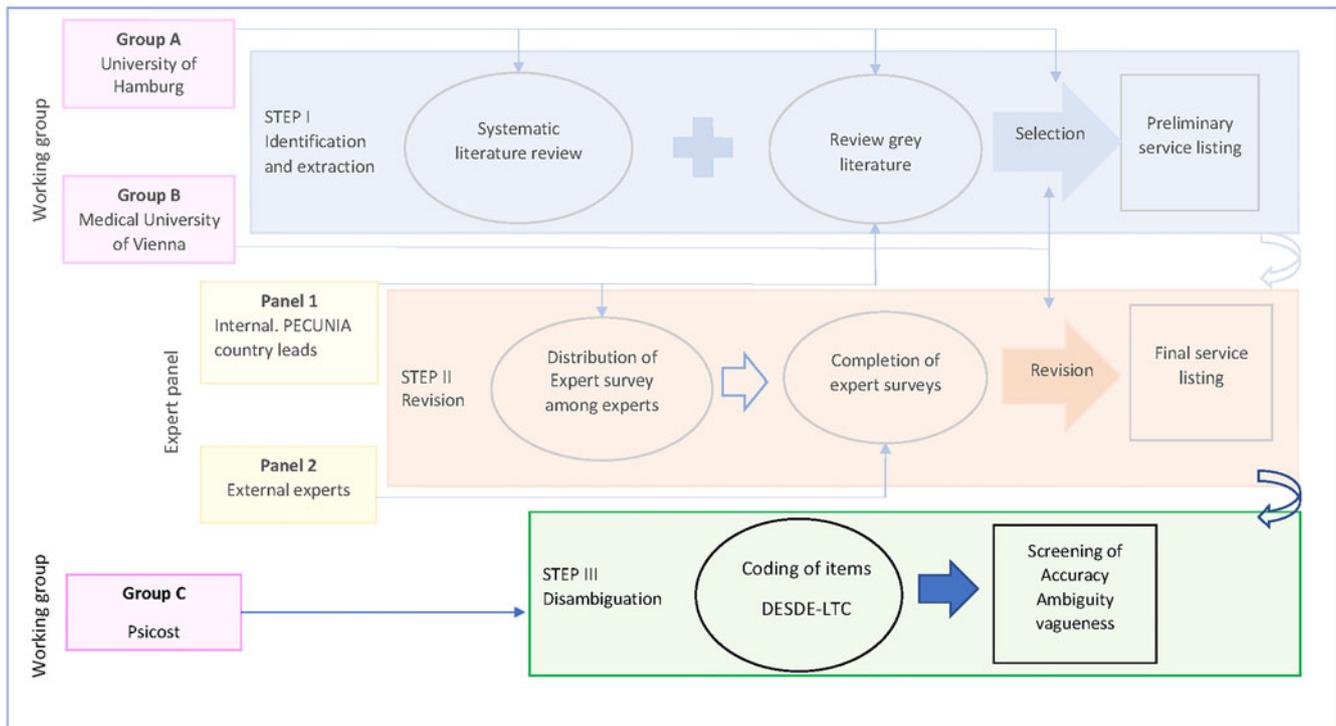
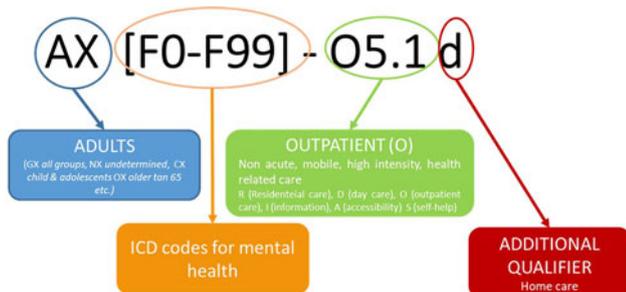**Fig. 1.** Multistep process for the ontoterminology study.

Fig. 2. Basic DESDE structure.

Results

Disambiguation of the final service listing

The basic listing included 35 terms corresponding to generic services for any health condition relevant to persons experiencing a mental disorder, and 21 that were specific services for mental health. This list was the basis for the master service list used for all next steps in the development of the PECUNIA costing tools. Initially only 13 terms (32%) had accurate definitions (not confusing, ambiguous or vague) according to the classification of reference. The disambiguation was analysed at three different levels:

Level 1: commensurability (the unit of analysis actually refers to services and not to other entities such as interventions or medical products)

Fifty terms (89%) were considered accurate at level 1 (online Supplementary Table 1). Six terms (11%) were considered confusing regarding the unit of analysis included in the definition. As an

example, 'emergency ambulance ride: a *special vehicle* used to take sick or injured people to a hospital or other health care facility in case of emergency', defined a device or medical good (a medicalised vehicle) but not the actual care provided. If this definition refers to the use of the vehicle, then it defines an intervention and not the ambulance service. The comparison between services targeting generic health (ICD-10) and mental health (F0-F9) showed no remarkable differences at level 1.

Level 2: target population clarity

The definition of the target population was considered accurate for 29 terms (52%). Three terms (6%) were judged ambiguous as they referred to two not-linked population groups (defined by age and/or diagnosis) at the same time, without explaining the service specificities for each of these groups. For example, 'nursing home: an inpatient care facility that offers care for *elderly OR disabled persons*'. Twenty-one terms (37%) were classified as vague as the definition of a population target was missing or it admitted too many possibilities (e.g. 'rehabilitation facility: a center or clinic where people recovering from *illness, injury or addiction* are treated'). Finally, three terms (5%) were considered confusing because they did not allow a clear classification of the target population, 'Sheltered housing for mentally ill persons: A sanctuary for temporary housing, set up to provide for the needs of *homeless people/women with mental disorders*, often including shelter, food, sanitation and other forms of support'. The meaning of 'homeless people/women with a mental disorder' is not clear and it did not match the name of the service (online Supplementary Table 1).

Level 3: service type clarity

The type of care provided by the service was judged accurate for the 23 terms (41%) that could be translated into a single MTC

code. Fifteen terms (27%) were rated as ambiguous because they needed two codes or admitted several code ranges. For example, 'polyclinic: a clinic that provides both general and specialist examinations and treatments' was coded as outpatient non-acute health-related care: O8.1-O10 (this range is used to address different intensities of frequency of care). However, it could also be coded as 'outpatient acute health-related care for a fixed number of hours': O4.1. Eleven terms (20%) were considered vague because a series of codes in different main care branches were necessary for classifying the term. Looking at the term 'rehabilitation facility: a center or clinic where people recovering from illness, injury or addiction are treated', the definition was so wide that the classification of the term required several codes from residential, day and outpatient care branches. Finally, six (11%) definitions were judged as confusing. For example, the two terms 'Outpatient healthcare at workplace: e.g. company physician, company nurse' and 'outpatient healthcare service at school: e.g. school physician, school nurse', included examples of professionals delivering the care but not an actual definition of the type of care provided. Additionally, two different names: 'vocational training' and 'individual vocational qualification' included the same definition 'Individual qualification training for a specific type of job' and therefore presented a problem of synonymity (online Supplementary Table 1). The six terms classified as confusing in level 1 (different unit of analysis) required expert interpretation in level 3 (coding service type), this was expressed in italics. The comparison between service targeting generic health (ICD-10) and mental health (F0-F9) showed remarkable differences regarding accuracy (22% *v.* 42%).

In total, 43 terms of the basic listing (77%) presented some kind of terminological inaccuracy.

The interrater reliability analysis showed statistically significant agreements for level 1 (k: 0.642, $p < 0.001$) and level 3 (k: 0.778, $p < 0.001$), while for level 2 (k: 0.875, $p < 0.001$) agreement was almost perfect. Agreement on the prototype DESDE codes was also high (k: 0.746, $p < 0.001$) (see online Supplementary material).

Discussion

This study aimed to identify the magnitude of the bias of non-commensurability and terminological unclarity bias in HSR and health economics by applying an international classification system to a set of services used by persons experiencing a mental health condition. The results are meant to be used for further processing of service terms for the development of the multi-national, multi-sectoral costing tools in the PECUNIA project. The approach was not comparing variation country by country but identifying an international basic listing of services relevant for mental health care. Despite an extensive process of revision prior to disambiguation, only 13 terms (23%) of the 56 were judged accurate. Eleven per cent of the terms in the final listing were not services according to the definition provided by the DESDE system and the related glossary of terms. In addition, 43 terms were unclear, and could not be used for international comparability. Nearly half of the terms lacked a clear definition of the target population and around 60% had problems in the definition of service types that impeded matching them to an MTC code even though we opted for broad categories within the MTC taxonomy to facilitate matching.

Our findings indicate that the terminology problem in HSR is extensive. Surprisingly enough, health economic guidelines

provide detailed information on the study designs, methods of analysis and interpretation of results but they do not mention this fundamental problem for regional or international comparability and for aggregation purposes (Simon, 2020). Similarly, the problem of service terminology is not even mentioned in international strategies that necessarily require comparison of service delivery such as the WHO Mental Health Gap Action Programme (mhGAP) (World Health Organization, 2008). A gap analysis cannot be conducted without a standardised description of local mental health services to allow aggregate comparisons of care systems across regions and countries. The approach presented in this paper contributes to minimise differentiation between services and to support financial incentives for effective health promotion and disease prevention. Health economic studies on services and their utilisation are key for RUM and cost calculation for efficiency (cost-effectiveness), equality (access and utilisation) and quality research.

Our experience replicated some of the findings described by Maciejewski *et al.* (2002) A detailed analysis of services in terms of cost measurement for economic evaluations reveals the necessity and usefulness of defining services using a coding system and taxonomical criteria rather than by 'text-based descriptions'.

Limitations

Firstly, the analysis of the terminology bias in healthcare is extremely challenging and may have problems with corroboration, even when we adapted a previously tested method (Maciejewski *et al.*, 2002) and used a standardised procedure. Secondly, the findings cannot be fully generalised to all areas of healthcare. We selected mental health care as case study due to its highest complexity of care provision stretching across numerous sectors (Salvador-Carulla *et al.*, 2006) including a mixture of health and social care services, the high variation and diversity in service provision (Johnson and Salvador-Carulla, 1998), and its high ambiguity in key aspects such as diagnosis (Keil *et al.*, 2016) and treatment interventions (Castelpietraa *et al.*, 2017; Castelpietra *et al.*, 2021). Third, we opted for a broad approach to disambiguation selecting the lower level of granularity in the MTC taxonomy and avoiding a detailed definition of the different subtypes of ambiguity and vagueness (Castelpietra *et al.*, 2021). The disambiguation data are related to one frame of reference (DESDE system) and cannot be generalised to other frames (e.g. Systems of Health Accounts 2.0, or SNOMED). However, the validity and the formal ontology conditions of the classification of services within these other frames have not been tested. Finally, we limited our analysis to English and did not account for the variation of terminology across other languages and contexts. In any case the reference tools ESMS and DESDE have been translated into Finnish, French, German, Italian, Polish, Portuguese, Norwegian and Spanish; and the reference coding system has been used across a wide variety of contexts in over 34 countries (Romero-Rodriguez-Alberca *et al.*, 2019).

Research and policy implications

Currently, the majority of comparative healthcare studies rely on official service names, without taking into account other key features of every service. Service health research, health economics, care gap analysis, quality and equality research should address terminological variability as a main source of systematic bias,

particularly, but not only, in bottom-up international comparative studies. For example, cost-effectiveness and comparative effectiveness research should compare the same units of analysis of service provision, and use a common vocabulary, which is feasible with a coding system such as the one provided by DESDE. This bias is also relevant in equity studies as equal access is a critical component of equity (Raine *et al.*, 2016).

Finally, an international glossary of service terms and a classification of service should be incorporated into the WHO International Family of Classifications as related classifications. Likewise, national classifications of services should provide an analysis of their semantic interoperability with international standards.

References

Alonso-Solís A, Ochoa S, Grasa E, Rubinstein K, Caspi A, Farkas K, Unoka Z, Usall J, Huerta-Ramos E, Isohanni M, Seppälä J, Reixach E, Berdún J, Corripio I, Alcalde F, D'amico E, Almazán C, Bitter I, Baccinelli W, Bonizzi C, Bulgheroni M, Mendivelso JC, Coenen T, Cohen A, Constant X, Escobar M, Fazekas K, Feldman Y, Gimenez E, van der Graaf S, Herman L, Hospedales M, Jääskeläinen E, Jewel C, Juola T, Jämsä T, Kaye R, Kokkinakis P, Koponen HJ, Marcó S, Mentzas G, Miettunen J, Moilanen J, Papas I, Paraskevopoulos F, Roldán A, Rubio-Abadal E, Sebú G, Seppälä A, Simonetti V, Stevens M, Tauro V, Triantafillou A, Unoka ZS, Vella V, Vermeir D and de Vita I (2020) A method to compare the delivery of psychiatric care for people with treatment-resistant schizophrenia. *International Journal of Environmental Research and Public Health* **17**, 7527.

Barnett PG (2009) An improved set of standards for finding cost for cost-effectiveness analysis. *Medical Care* **47**, S82–S88.

Castelpietraa G, Salvador-Carulla LE, Almborg AH, Fernandez A and Madden R (2017) Working draft: classifications of interventions in mental health care. An expert review. *The European Journal of Psychiatry* **31**, 127–144.

Castelpietra G, Simon J, Gutierrez-Colosía MR, Rosenberg S and Salvador-Carulla L (2021) Disambiguation of psychotherapy – a search for meaning. *British Journal of Psychiatry* **219**, 532–537.

Fischer C, Mayer S, Perić N, Simon J and PECUNIA Group (2022) Establishing a comprehensive list of mental health-related services and resource use items in Austria: a national-level, cross-sectoral country report for the PECUNIA project. *PLoS ONE* **21**, e026209.

Fortune N, Madden R and Almborg AH (2018) Use of a new international classification of health interventions for capturing information on health interventions relevant to people with disabilities. *International Journal of Environmental Research and Public Health* **15**, 145.

García-Alonso CR, Almeda N, Salinas-Pérez JA, Gutiérrez-Colosía MR, Uriarte-Uriarte JJ and Salvador-Carulla L (2019) A decision support system for assessing management interventions in a mental health ecosystem: the case of Bizkaia (Basque Country, Spain). *PLoS ONE* **14**, e0212179.

Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J and Loder E (2013) Consolidated health economic evaluation reporting standards (cheers) statement. *International Journal of Technology Assessment in Health Care* **29**, 117–122.

Iglehart JK (2009) Prioritizing comparative-effectiveness research – IOM recommendations. *New England Journal of Medicine* **361**, 325–328.

Johnson S and Salvador-Carulla L (1998) Description and classification of mental health services: a European perspective. *European Psychiatry: The Journal of the Association of European Psychiatrists* **13**, 333–341.

Keil G, Keuck L and Hauswald R (2016) *Vagueness in Psychiatry*. Oxford: Oxford University Press.

Landis JR and Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.

Maciejewski ML, Diehr P, Smith MA and Hebert P (2002) Common methodological terms in health services research and their symptoms. *Medical Care* **40**, 477–484.

Mayer S, Fischer C, Zechmeister-Koss I, Ostermann H and Simon J (2020) Are unit costs the same? A case study comparing different valuation methods for unit cost calculation of general practitioner consultations. *Value in Health* **23**, 1142–1148.

Mayer S, Berger M, Konnopka A, Brodzsky V, Evers S, Hakkaart-van Roijen L, Gutiérrez-Colosía MR, Salvador-Carulla L, Park A, Hollingworth W, García-Pérez L, Simon J and on behalf of the PECUNIA group (2022) In search for comparability: the PECUNIA reference unit costs for health and social care services in Europe. *International Journal of Environmental Research and Public Health* **19**, 3500.

Montagni I, Salvador-Carulla L, Mcdaid D, Straßmayr C, Endel F, Näätänen P, Kalseth J, Kalseth B, Matosevic T, Donisi V, Chevreul K, Prigent A, Sfectu R, Pauna C, Gutiérrez-Colosia MR, Amaddeo F and Katschnig H (2018) The REFINEMENT glossary of terms: an international terminology for mental health systems assessment. *Administration and Policy in Mental Health and Mental Health Services Research* **45**, 342–351.

Moreno K, Sanchez E and Salvador-Carulla L (2008) Methodological advances in unit cost calculation of psychiatric residential care in Spain. *The Journal of Mental Health Policy and Economics* **11**, 79–88.

Noben CY, De Rijk A, Nijhuis F, Kottner J and Evers S (2016) The exchangeability of self-reports and administrative health care resource use

measurements: assessment of the methodological reporting quality. *Journal of Clinical Epidemiology* **74**, 93–106.

OECD Eurostat WHO (2017) *A System of Health Accounts 2011: Revised Edition*. Paris: OECD publishing https://www.oecd-ilibrary.org/social-issues-migration-health/a-system-of-health-accounts-2011_9789264270985-en.

Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F and Stevens D (2016) SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Quality and Safety* **25**, 986–992.

O'Neill K, Takane M, Sheffel A, Abou-Zahr C and Boerma T (2013) Monitoring service delivery for universal health coverage: the service availability and readiness assessment. *Bulletin of the World Health Organization* **91**, 923–931.

Raine R, Fitzpatrick R, Barratt H, Bevan G, Black N, Boaden R, Bower P, Campbell M, Denis JJ, Devers K, Dixon-Woods M, Fallowfield L, Forder J, Foy R, Freemantle N, Fulop NJ, Gibbons E, Gillies C, Goulding L, Grieve R, Grimshaw J, Howarth E, Lilford RJ, McDonald R, Moore G, Moore L, Newhouse R, O'Cathain A, Or Z, Papoutsi C, Prady S, Rycroft-Malone J, Sekhon J, Turner S, Watson SI and Zwarenstein M (2016) Challenges, solutions and future directions in the evaluation of service innovations in health care and public health. *Health Services and Delivery Research* **4**, 1–136.

Roche C (2012) Ontoterminology: how to unify terminology and ontology into a single paradigm, in Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC. Available at http://www.lrec-conf.org/proceedings/lrec2012/pdf/567_Paper.pdf.

Romero-Lopez-Alberca C, Gutierrez-Colosia MR, Salinas-Pérez JA, Almeda N, Furst M, Johnson S and Salvador-Carulla L (2019) Standardised description of health and social care: a systematic review of use of the ESMS/DESDE (European Service Mapping Schedule / Description and Evaluation of Services and DirectoriEs). *European Psychiatry* **61**, 97–110.

Rosen A, Rock D and Salvador-Carulla L (2020) The interpretation of beds: more bedtime stories, or maybe they're dreaming? *Australian and New Zealand Journal of Psychiatry* **54**, 1154–1156.

Salinas-Pérez JA, Gutiérrez-Colosía MR, Romero López-Alberca C, Poole M, Rodero-Cosano ML, García-Alonso CR and Salvador-Carulla L (2020) Todo está en el mapa: Atlas Integrales de Salud Mental como herramientas para la planificación de servicios de salud mental. *Gaceta Sanitaria* **34**, 11–19.

Salvador-Carulla L and Hernández-Peña P (2011) Economic context analysis in mental health care. Usability of health financing and cost of illness

studies for international comparisons. *Epidemiology and Psychiatric Sciences* **20**, 19–27.

Salvador-Carulla L, Atienza C, Romero C and PSICOST Group (1999) Use of the EPCAT model for standard description of psychiatric services: the experience in Spain. In Guimón J and Sartorius N (eds), *Manage or Perish?. The Challenges of Managed Mental Health Care in Europe*. New York: Kluwer Academic/Plenum Publishers, pp. 359–368.

Salvador-Carulla L, Haro JM and Ayuso-Mateos JL (2006) A framework for evidence-based mental health care and policy. *Acta Psychiatrica Scandinavica* **111**, 5–11.

Salvador-Carulla L, Alvarez-Galvez J, Romero C, Gutiérrez-Colosía MR, Weber G, McDaid D, Dimitrov H, Sprah L, Kalseth B, Tibaldi G, Salinas-Perez JA, Lagares-Franco C, Romá-Ferri MT and Johnson S (2013) Evaluation of an integrated system for classification, assessment and comparison of services for long-term care in Europe: the eDESDE-LTC study. *BMC Health Services Research* **13**, 218.

Salvador-Carulla L, Amaddeo F, Gutiérrez-Colosía MR, Salazzari D, Gonzalez-Caballero JL, Montagni I, Tedeschi F, Cetrano G, Chevreul K, Kalseth J, Hagmair G, Straßmayr C, Park A, Sfetcu R, Wahlbeck K and Garcia-Alonso C (2015) Developing a tool for mapping adult mental health care provision in Europe: the REMAST research protocol and its contribution to better integrated care. *International Journal of Integrated Care* **15**.

Simon J (2020) Health economic analysis of service provision. In Geddes JR, Andreasen NC and Goodwin GM (eds), *New Oxford Textbook of Psychiatry*, 3rd Edn. Oxford: Oxford University Press, pp. 1384–1391.

The PECUNIA group (2018) *Consortium*. Available at https://www.pecunia-project.eu/about/consortium.

Thorn JC, Coast J, Cohen D, Hollingworth W, Knapp M, Noble SM, Ridyard C, Wordsworth S and Hughes D (2013) Resource-use measurement based on patient recall: issues and challenges for economic evaluation. *Applied Health Economics and Health Policy* **11**, 155–161.

van Lier LI, Bosmans JE, van Hout HPJ, Mokkink LB, van den Hout WB, de Wit GA, Dirksen CD, Nies HLGR, Hertogh CMPM and van der Roest HG (2018) Consensus-based cross-European recommendations for the identification, measurement and valuation of costs in health economic evaluations: a European Delphi study. *European Journal of Health Economics* **19**, 993–1008.

World Health Organization (2008) *Mental Health Gap Action Programme – Scaling Up Care for Mental, Neurological, and Substance Use Disorders*. World Health Organization. Available at https://apps.who.int/iris/handle/10665/43809.