An international comparison of efficiency of inpatient mental health care systems

Valerie Moran, Rowena Jacobs

Centre for Health Economics, University of York, Alcuin A Block, Heslington, York YO10 5DD, United Kingdom

ARTICLE INFO

Article history:
Received 25 October 2012
Received in revised form 11 June 2013
Accepted 19 June 2013

Keywords:
Cross-country comparisons
Inpatient mental health services
Data envelopment analysis
Cluster analysis
Bootstrapping

ABSTRACT

There is a fundamental gap in the evidence base on quantitative cross-country comparisons of mental healthcare systems due to the challenges of comparative analysis in mental health including a paucity of good quality data. We explore whether existing limited data sources can potentially be exploited to examine technical efficiency of inpatient mental healthcare systems in 32 OECD countries in 2010. We use two analytical approaches: Data Envelopment Analysis (DEA) with bootstrapping to produce confidence intervals of efficiency scores and country rankings, and Cluster Analysis to group countries according to two broad efficiency groupings. We incorporate environmental variables using a two-stage truncated regression. We find slightly tighter confidence intervals for the less efficient countries which loosely corresponds with the ‘inefficient’ cluster grouping in the Cluster Analysis. However there is little stability in country rankings making it difficult with current data to draw any policy inferences. Environmental factors do not appear to significantly impact on efficiency scores. The most pressing pursuit remains the search for better national data in mental healthcare to underpin future analyses. Otherwise the use of any sophisticated analytic techniques will prove futile for establishing robust conclusions regarding international comparisons of the performance of mental healthcare systems.

1. Introduction

The World Health Organization (WHO) World Health Report (WHR), while provoking much critical debate at a conceptual and empirical level, made an important contribution in seeking to provide a quantitative assessment of comparative health system performance, bringing the topic to the attention of policy makers worldwide [1]. Quantitative cross-country comparisons have been used in many different contexts and while they present many challenges [2], can constitute a rich source of evidence for policy makers [3]. There has been a substantial effort to conduct such quantitative cross-country comparisons for healthcare systems focusing on physical health or particular disease conditions [4,5]. Very little research endeavour has focused on cross-country comparisons of mental health care performance. One of the key reasons for this has been due to substantial gaps in the data for mental health compared to physical health [6]. There is therefore a fundamental gap in the evidence base on quantitative comparative mental health care system performance.

Some international efforts have been made to establish cross-national comparative benchmarking for mental health indicators. The National Institute for Health and Welfare in Finland together with the European Commission Health Monitoring Programme developed a set of...
mental health performance indicators for European Union countries [7]. The Organization for Economic Cooperation and Development (OECD) has also identified key mental health quality indicators for use in international benchmarking [8]. The OECD Health Care Quality Indicators (HCQI) project currently collects two mental health indicators – re-admission rates for schizophrenia and bipolar disorders [9], while the WHO have developed the Assessment Instrument for Mental Health Systems (WHO-AIMS 2.2) to assess mental health system performance on a multi-item scale [10]. The WHO’s Department of Mental Health and Substance Abuse has designed ‘Project Atlas’ to collect data on resource use for mental health on a global basis [11].

There are particular challenges to mental health comparative analysis. Lauriks et al. [12] report on many national initiatives to develop performance measures for mental health. They provide a systematic review identifying 1480 unique performance indicators that are used internationally to measure the performance of public mental health care. They find that less than 3% of performance indicators actually assess the efficiency, cost or expenditure of mental health care systems. Most countries which collect data on mental health quality and performance, tend to focus on hospital care and measures of utilization [13]. There is also a wide variation between countries in the indicators which are collected, reflecting a focus on local priorities. This makes international comparative work more difficult. Efficiency and performance measurement is also immensely more challenging in mental than in somatic health care due to difficulties in measuring outputs and outcomes, the complex nature of mental health care [14], interactions with non-health sectors and the seriously marginalizing social consequences of mental ill health which may impact on measurement [15].

There is a substantial gap between the burden caused by mental disorders and the resources available to prevent and treat them. Neuropsychiatric disorders are estimated to account for 14% of the global burden of disease [16,17]. The economic costs to societies of mental health problems are enormous, including lost employment, absenteeism and sick leave, reduced performance at work, lost leisure opportunities and premature mortality [18]. Conservatively the costs of poor mental health are estimated to account for 3–4% of GDP in the European Union (EU) alone, yet nowhere in the EU does spending on mental health much exceed 1% of GDP [19]. Funding for mental health as a proportion of the total health budget in the EU ranges from around 14% in England to much less than 4% in some countries including Bulgaria, the Czech Republic, Poland and Portugal [15]. The relatively low level of resources allocated to mental health and the commensurate large burden of disease, makes the efficient use of resources imperative. Moreover, there is an increased impetus to this need given the current economic context which has seen a consolidation in overall health (as well as mental health) budgets in a number of OECD countries [20].

We aim to examine the feasibility of quantitative cross-country comparison of mental healthcare systems by exploiting available, though limited data sources on mental health care systems. We examine whether different analytical approaches are able to offer insights into quantitative international comparison in this much neglected area. Our research question asks whether it is feasible to use different methodological approaches on current data to analyze the technical efficiency of inpatient mental health care in a cross-country context. Technical efficiency refers to the extent to which a country secures the maximum output attainable given its inputs [21]. While many OECD countries are increasingly moving towards community-based models of mental health care, inpatient care still constitutes a core component of care and one which accounts for a significant amount of resources. On a global basis, 67% of mental health expenditures are directed towards mental hospitals (54% in high income (World Bank classification) countries and 60% in the EURO region) [11].

We try to tease out the relative efficiency of inpatient mental health care systems across 32 OECD countries. We propose the use of two analytical approaches, Data Envelopment Analysis (DEA) with bootstrapping [22] and Cluster Analysis [23] which may offer potential for exploring the data. We examine whether these techniques provide complementary results in terms of distinguishing groups of countries that are high or low performers in terms of efficiency. Both DEA and Cluster Analysis are explorative data mining techniques and we examine whether these methods applied to existing, even very limited data sources, can offer any scope for cross-country comparative insights. DEA has previously been applied to measure the efficiency of mental health services primarily in national contexts [24–26], but not in a cross-country context. There have been very few cross-country comparative studies of health systems using Cluster Analysis [27].

The Scientific Peer Review Group that commented on the WHO (2000) WHR [1], suggested future comparative efficiency analyses should explore exogenous factors that may impact on health system performance [28]. We include four environmental variables in the efficiency analysis to examine whether they contribute to efficiency estimates and whether they impact on countries’ ability to improve their efficiency. We also employ a two-stage DEA analysis, where the bootstrap DEA scores are regressed against a set of environmental variables using a truncated regression analysis, to assess the impact of potential exogenous factors on technical efficiency [29,30].

2. Data

We use data from the 2012 edition of the OECD Health Database and acknowledge there are a number of significant limitations around data availability and measurement [9]. The majority of the data covers the year 2010 but there are a notable number of exceptions where data from earlier years had to be used in order to get the most complete dataset possible. This is clearly far from ideal since there may be an intervening period of a number of years between countries in which significant change may have occurred in the mental healthcare systems. Yet countries are being evaluated as if all inputs and outputs are occurring contemporaneously. Furthermore, there are important limitations to cross-section data – they could lead to misleading inferences if variables are influenced by systematic
fluctuations, and there is no way of accounting for omitted explanatory variables or controlling for unobservable heterogeneity [31]. However, it was deemed important to try to get some data coverage rather than none at all at a single point in time for as many countries as possible at the very least, to start to get a snapshot of what is happening in terms of inpatient mental health care efficiency.

As mentioned, very few performance indicators assess measures of efficiency, cost or expenditure of mental health care systems. It is recognized that information on health care costs and health expenditure is often limited in scope and comparability, and healthcare expenditure data by disease are often particularly difficult to obtain for research purposes [32]. While we explored the use of mental health care system expenditure data from the WHO Atlas [11], the coverage was too low to allow analysis with a sufficient sample size. Ideally, we would seek to have data on health outcomes, however OECD Health Data [9] on suicide rates, anti-depressant consumption, and readmission rates for schizophrenia and bipolar disorder, had to be discarded due to low coverage. Equally we would want data on case-mix adjustment, measures of quality, responsiveness, and experience, to provide a fuller picture of inpatient mental health care efficiency, but were constrained by data availability. Nevertheless, we seek to explore the feasibility of using existing data sources and exploiting them to the full to assess their ability in providing insights when scrutinized under different methodological approaches.

Table 1 describes the variables used in the DEA analysis along with descriptive statistics, the mean and standard deviation. We include a number of different inputs to inpatient mental healthcare services (psychiatrists, beds, length of stay) and outputs (discharges). We also control for potential environmental factors that may impact on countries’ abilities to transform inputs into outputs, such as key population socio-economic factors. We describe each variable in turn.

Supplemental Fig. S1 shows a cross-country comparison of the number of psychiatrists per 1000 population and highlights that Switzerland is an outlier relative to other OECD countries. This may partially be explained by a relatively high annual average growth rate of psychiatrists of 5.4% between 2000 and 2010. It is also notable that Chile and Mexico have very low numbers of psychiatrists, a reflection of the overall low rates of medical doctors [9].

As also shown in Fig. S1, Belgium and Netherlands have a relatively high number of psychiatric care beds and this may provide hospitals with incentives to keep patients longer. Measures of the availability of psychiatric beds may vary substantially across countries because of the variation in the organization and management of mental health services within health care systems. The evidence base to support a community care centred approach has continued to grow [33,34]. This changing evidence base means that countries are now at very different stages in rebalancing their mental health systems, so as to make community based care the mainstay of the system [11]. Across nearly all of western Europe there has been a shift in the balance of care with the closure of many psychiatric hospitals and the transfer of beds to general hospitals. In much of northern Europe this has been accompanied by investment in social and community care based services. Elsewhere, in Mediterranean countries such as Italy, Portugal and Spain, there has been little investment in community-based alternatives and much of the responsibility for support rests with families. In contrast, in much of central and eastern Europe long-stay very large and often isolated psychiatric hospitals and social care homes (internats) still dominate; there are few incentives to change the balance of care, particularly where local communities rely on these institutions for employment.

Average length of stay (ALoS) is often used as a measure of efficiency or a surrogate measure for cost or resource use. All other things being equal, a shorter stay will reduce the cost per discharge and shift care from inpatient to less expensive post-acute settings [3]. However ALoS is also often seen as an important quality indicator [35] which may be associated with health outcomes, recovery and readmission rates [36], making it a valid input in the DEA approach [37]. Fig. S1 of the supplemental data shows a cross-country comparison of ALoS for OECD countries. It highlights the two countries with outlier values on ALoS (Korea and Greece). Financial incentives inherent in hospital payment methods can influence length of stay. In particular, Korea has a per diem payment system for inpatient care which may incentivize lengths of stay longer than are clinically necessary.

We would ideally seek to include health outcomes alongside health output measures such as discharges, since health gain is the key indicator of the success of a health system, however analysts are often constrained in practice to examine efficiency on the basis of measures of activities such as discharges, as proxies for health outcomes. Furthermore, mental health outcomes are notoriously difficult to capture [15]. Measuring activities can address a fundamental shortcoming of outcome measurement which is how much of the outcome is attributable to the health care system [21]. The final diagram in Fig. S1 shows a cross-country comparison of discharges for mental and behavioural disorders per 1000 population. Germany, Austria and Finland have the highest rate of discharges while Israel and Mexico have the lowest. In general, those countries that have more hospital beds tend to have higher discharge rates, thus it is not unusual for Mexico to have the lowest rate of discharges given that it also has the lowest rate of beds.

Alcohol consumption is included as an environmental factor since it is particularly strongly associated with mental health problems [38]. OECD Health Data contains data on the annual consumption of pure alcohol in litres per person aged 15 years and above. Average per capita consumption may fail to account for a particularly dangerous pattern of consumption, namely large quantities of alcohol at a single session (“binge drinking”) which is on the rise in some countries and social groups. Nevertheless, population consumption patterns may be an important environmental factor because rates of co-occurring mental and substance misuse conditions have been documented with an estimated 42.7% of adults in the US aged 15–54 with an alcohol or drug disorder also having a mental disorder and 14.7% of those with a mental disorder also having an alcohol or drug disorder [38].
Table 1
Description of data used in efficiency analysis and model specifications.

| Variable                  | Year 2010 except | Definition                                                                                                                                                                                                                                                                                                                                 | Mean (std dev) | M1 | M2 | M3 | M4 |
|---------------------------|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|----|----|----|----|
| **Inputs**                |                  |                                                                                                                                                                                                                                                                                                                                          |                |    |    |    |    |
| Psychiatrists             | 2007 (Slovak Republic) 2009 (Australia, Denmark, Italy, Netherlands, Sweden) | Per 1000 population. The figures normally include psychiatrists, neuropsychiatrists and child psychiatrists. Psychologists are excluded. The numbers are presented as head counts, regardless of whether psychiatrists work full-time or part-time. | 0.15 (0.08)    | ✓  | ✓  | ✓  | ✓  |
| Psychiatric beds         | 2008 (Australia, Canada, Greece, Netherlands, USA) | Per 1000 population. Psychiatric care beds are beds accommodating patients with mental health problems. They include beds in psychiatric departments of general hospitals, and all beds in mental health and substance abuse hospitals. | 0.62 (0.37)    | ✓  | ✓  | ✓  | ✓  |
| Average length of stay   | 2007 (Greece) 2008 (Belgium) 2009 (Australia, Canada, Chile, Hungary, Italy, Netherlands, Portugal, USA) | Average length of stay (ALoS) refers to the average number of days that patients with a primary diagnosis of a mental or behavioural disorder (ICD-10 F00–F99 and ICD-9 290–319) spend in hospital. It is measured by dividing the total number of days stayed by all inpatients during a year by the number of admissions or discharges. Day cases are excluded. | 24.32 (23.20) | ✓  | ✓  | ✓  | ✓  |
| **Outputs**               |                  |                                                                                                                                                                                                                                                                                                                                          |                |    |    |    |    |
| Discharges                | 2007 (Greece) 2008 (Belgium) 2009 (Australia, Canada, Chile, Hungary, Italy, Netherlands, Portugal, USA) | Per 1000 population. Hospital discharge is defined as the release of a patient with a primary diagnosis of a mental or behavioural disorder (ICD-10 F00–F99 and ICD-9 290–319) who has stayed at least one night in hospital. It includes deaths in hospital following inpatient care. Same-day discharges are usually excluded, with the exceptions of Chile, France, Korea, Norway, Poland, the Slovak Republic, Turkey and the United States which include some same-day separations. | 543.18 (428.08) | ✓  | ✓  | ✓  | ✓  |
| **Environmental variables** |                  |                                                                                                                                                                                                                                                                                                                                          |                |    |    |    |    |
| Alcohol consumption      | 2007 (Israel, Korea, Portugal) 2008 (Belgium, Luxembourg, Mexico) 2009 (Australia, Austria, Czech Republic, Germany, Greece, Hungary, Italy, Netherlands, Slovak Republic, Spain, USA) | Alcohol consumption is defined as annual sales of pure alcohol in litres per person aged 15 years and over. | 9.63 (2.79)    | ✓  | ✓  | ✓  | ✓  |
Table 1 (Continued)

| Variable        | Year 2010 except | Definition                                                                                                           | Mean (std dev)                      | M1 | M2 | M3 | M4 |
|-----------------|------------------|---------------------------------------------------------------------------------------------------------------------|------------------------------------|----|----|----|----|
| Income          | –                | Income is defined as gross domestic product (GDP) per 1000 population measured in USSPPP. Purchasing power parities (PPPs) are the rates of currency conversion that eliminate the differences in price levels between countries. Per capita volume indices based on PPP converted data reflect only differences in the volume of goods and services produced. Comparative price levels are defined as the ratios of PPPs to exchange rates. They provide measures of the differences in price levels between countries. The PPPs are given in national currency units per US dollar. The price levels and volume indices derived using these PPPs have been rebased on the OECD average. | 33,977,000 (13,861,000)            |    |    | ✓  | ✓  |
| Education       | –                | Education is defined in terms of the percentage of the population with at least upper secondary education. Income and education are included in the model as proxies of socio-economic status. Numerous studies have documented the association between the prevalence of mental disorders and lower socio-economic status [39,40]. This relationship may be due to (1) social causality, whereby social and economic circumstances determine mental health outcome; or (2) social selection in that people with mental illness descend the social scale because of their illness (or conversely cannot rise up the social scale) and their mental illness is caused by factors other than socio-economic ones. | 73.53 (16.75)                      |    |    | ✓  |    |
| Unemployment    | –                | Unemployment is measured as a percentage of the civilian labour force. | 8.62 (3.86)                        |    |    | ✓  | ✓  |

Source: OECD (2012) OECD Health Data 2012, OECD Publishing, Paris. OECD (2011) Education at a Glance 2011, OECD Publishing, Paris. OECD (2012), OECD Main Economic Indicators, Paris.

Note: N = 32.
M1 = Model 1, M2 = Model 2, M3 = Model 3, M4 = Model 4.

3. Methods

Data Envelopment Analysis (DEA) is a non-parametric approach that uses the simple notion that a mental healthcare system that produces more output than another for the same amount of input can be considered more efficient [41]. DEA makes minimal assumptions about the underlying technology making it less susceptible to specification error, but with no scope for random error [42–44]. The choice of DEA to measure efficiency in this study can be justified by its relative strength, along with bootstrapping procedures [45], in dealing with small sample sizes [21]. DEA has been criticized for being deterministic, vulnerable to sampling noise in the efficiency estimates and sensitive when outliers exist [46], particularly when sample sizes are small as in our study. Hence, the bootstrapping approach [45] is employed to address some of these limitations.

The construction of the efficiency frontier is based on 'best observed practice' and is only an estimate of the true, unobserved efficiency frontier. Since statistical estimators of the frontier are obtained from finite samples, the corresponding measures of efficiency are sensitive to the sampling variations of the obtained frontier [47]. Bootstrapping is a statistical tool for analyzing the sensitivity of measured efficiency scores to sampling variation. We use a smoothed bootstrap whereby independent and identically distributed bootstrap samples are drawn from a smooth consistent estimator of the joint density of the production
set of physically attainable points [45]. We repeatedly simulate the data-generating process through re-sampling and apply the original estimator to each simulated sample so that the resulting estimates mimic the sampling distribution of the original estimator. We apply the bootstrap algorithm applied by Simar and Wilson [47]. While it remains true that DEA does still not allow for random noise or measurement error in the data, the bootstrap procedure addresses some of the criticism by correcting the sampling bias and allows for classical statistical inference, thus confidence intervals can be generated for the efficiency estimates and rankings [45,47]. One can then examine the sensitivity of DEA efficiency scores and country rankings to variations in sample composition. Confidence intervals for countries on the fringes of the data may be wider, suggesting the estimates are based on thin data, and should be interpreted cautiously. The reliability of an efficiency score and resultant ranking depends on the density of observations in the region of the frontier where a country is located. Countries with atypical inputs and outputs tend to be considered efficient but this result may simply be a consequence of the lack of comparable observations. We therefore test the sensitivity of our results by running numerous different models and provide confidence statements around the point estimates [48,49].

We employ variable returns to scale (VRS) [50] since we have ratio data [51] and an input orientation because many healthcare systems may face resource constraints and incentives may be more readily oriented towards minimizing inputs such as length of stay or beds.

We analyze a baseline model (Model 1) with no environmental adjusters and test the sensitivity to the inclusion of various environmental variables. The four environmental variables enter the model as uncontrolled inputs [52,53]. Table 1 specifies the variables used in 4 models which illustrate the inclusion of the key characteristics of the inpatient mental healthcare system and various combinations of environmental variables. All 4 models include the 3 key inputs and 1 output, with Model 1 including no environmental adjusters. Model 2 includes only alcohol consumption which is considered the environmental variable most closely aligned to mental health problems. Model 3 tests the inclusion of a key socio-economic determinant, income, along with a key measure of labour market participation, unemployment. Model 4 includes all environmental adjusters.

An alternative approach to dealing with environmental variables is a two-stage approach whereby DEA scores in the first stage are regressed on covariates in the second stage using a tobit regression. The DEA efficiency estimates are however seriously correlated, invalidating classical statistical inference. We use the approach proposed by Simar and Wilson [22] which is an alternative second stage truncated regression estimation procedure using the bootstrap estimators to solve the dependency problem. The bias-corrected efficiency scores from the first stage are therefore regressed on the four environmental factors in the second stage. If significant, variables with an estimated positive coefficient would be interpreted as having a positive impact on efficiency in the truncated regression. We test for multi-collinearity of the environmental covariates using the Variance Inflation Factor (VIF).

Cluster Analysis is a technique used to group similar variables together into distinct categories. Most applications of Cluster Analysis partition the data so that each object or observation belongs to a single cluster and the complete set of clusters contain all of the data [23]. There are various ways in which clusters can be formed and we utilize hierarchical agglomerative clustering whereby each country begins as a unique cluster and is consequently grouped with other countries to create smaller clusters so that the average distance between all countries in the resulting cluster is minimized [23]. We test a number of different algorithms (complete linkage, Ward’s method, centroid method and median method) and base our final Cluster Analysis on the average linkage algorithm to connect countries to form clusters based on their distance. Euclidean distances are applied as these are most commonly used for ratio data [54]. We again use a number of different models to test the robustness of the results, but present results for Model 1, our baseline model for the inpatient mental healthcare system.

At different distances, different clusters will form, which can be represented using a dendrogram or tree diagram which provides a visual display of the clusters. The dendrogram shows the hierarchical clustering which the algorithm provides. The hierarchy of clusters merge with each other at certain distances. In a dendrogram, the y-axis marks the distance at which the clusters merge, or the dissimilarity measure, while the countries are placed along the x-axis such that the clusters don’t mix.

For the DEA analyses we use the FEAR [55] (Version 1.15) package in R (version 2.13.1). The Cluster Analysis and the truncated regression were performed in Stata 12.1. The data and the R and Stata code are available as (online) appendices.

4. Results

Fig. 1 shows the results of the 4 DEA models. Countries are ranked from highest to lowest efficiency scores and 95% confidence intervals surround the point estimates. The figure also shows the country rankings and ranking distributions bounded at the 5th and 95th percentiles of the distribution for each model. It is interesting to note that for some countries the ranking falls outside the 95% interval. This is because the distributions of DEA efficiency scores are typically skewed [47].

Countries with higher efficiency estimates like Slovenia, Korea, Poland, Denmark, Hungary, Italy, Chile, United States, Austria, Norway, Turkey, Mexico, and Germany, tend to have much wider confidence intervals suggesting less security around the point estimates. These countries do however tend to have higher efficiency scores regardless of which model is used. There are however countries which are highly sensitive to model specification, such as Switzerland. This is because Switzerland is an outlier on some dimensions; it has very high numbers of psychiatrists (see Fig. S1), but also has high levels of beds and discharges, and low unemployment.
Fig. 1. The 4 DEA model results and country-specific rankings.
Fig. 1. (Continued).
Table 2
Truncated regression results for environmental variables.

| Variable       | Coefficient (std error) | 95% confidence interval |
|----------------|-------------------------|-------------------------|
|                |                         | Lower bound | Upper bound |
| Constant       | 0.944 (0.033)***        | 0.469       | 1.420       |
| Alcohol consumption | −0.002 (0.016)    | −0.033      | 0.030       |
| Income         | −0.005 (0.000)         | −0.011      | 0.002       |
| Education      | 0.001 (0.003)          | −0.004      | 0.006       |
| Unemployment   | −0.015 (0.011)         | −0.037      | 0.007       |

*** Significant at 1% level; total number of iterations = 2000.

There appears to be greater stability for the very bottom ranked countries with tighter confidence limits. United Kingdom, Greece and Netherlands, tend to consistently appear towards the bottom of the league table.

Our baseline Model 1 without any environmental variables appears to have slightly narrower confidence intervals compared to a model which incorporates all potential environmental variables, Model 4. This suggests the results for the more parsimonious model focusing just on the inpatient mental healthcare system may offer slightly more secure efficiency estimates.

When, however, we examine the ranking distributions for countries in Fig. 1, the confidence intervals are very much wider and it is much more difficult to make any definitive statements about relative comparisons of country standings. Indeed given the wide overlapping intervals, the rankings for the majority of countries appear somewhat arbitrary. It is however the case that countries with overlapping confidence intervals from the efficiency scores in the first part of Fig. 1, which should in principle share rankings since they are indistinguishable, are forced to have separate rankings from 1 to 32. This enforces to some extent more variability in the rankings compared to the efficiency scores, resulting in wider confidence intervals in the second part of Fig. 1.

Nevertheless, the results suggest a small number of countries, Greece, United Kingdom and Netherlands, tend to remain clustered at the bottom of the rankings with marginally tighter confidence intervals, suggesting they may be deserving of further scrutiny to ascertain the reasons why their apparent relative performance may fall short of their counterparts. Greece and the United Kingdom for instance have relatively longer lengths of stay and the Netherlands has low rates of discharges.

Overall, the DEA results suggest we are able to discern some broad differences in patterns of efficiency between countries, but we should proceed with extreme caution when it comes to any kind of ranking.

Truncated regression results for the environmental variables are presented in Table 2. We find no multi-collinearity between regressors (VIF = 1.32). None of the environmental variables are significant in the model suggesting that they do not contribute to explaining variation in cross-country efficiency scores. This gives some credence to the appropriateness of the baseline Model 1, given that external factors are not playing a key factor in the efficiency analysis.

Fig. 2 presents the dendrogram showing two clusters of countries, the ‘efficient’ cluster with 12 countries and the ‘inefficient’ cluster with 20 countries. The average efficiency within the 2 clusters are 0.71 (C.I. 0.63–0.8) and 0.67 (C.I. 0.58–0.77) respectively though the difference is not statistically significant, suggesting the apparent differences between the two country groupings is purely descriptive. Nevertheless, the Cluster Analysis results, loosely correlate with the results from the DEA analysis in Model 1. The ‘efficient’ countries with the wider confidence intervals tend to map more poorly to the dendrogram and Korea, Chile, Turkey, Denmark, Norway, Italy, Slovenia, and Mexico appear to be clustered ‘erroneously’ in the ‘inefficient’ group in the Cluster Analysis. The set of inefficient countries in the DEA analysis with the narrower confidence intervals all correspond with the inefficient cluster grouping in the Cluster Analysis and appear to have been categorized in a consistent way with the DEA model. Though the results are very tentative given the poor data, the correspondence in groupings given by the two methods cautiously suggests the potential for the two analytical techniques to potentially discriminate between groups of countries in terms of efficiency if better data were available.

5. Discussion
This paper aims to examine the feasibility of quantitative cross-country comparison in mental health by exploiting available OECD data on inpatient mental healthcare services using cross-section data on 32 countries. We examine the use of two analytical approaches to explore relative efficiency on available, albeit very limited data, namely DEA with bootstrapping and Cluster Analysis. We examine whether these methods produce complementary results in categorizing countries in terms of broad patterns of efficiency. We use inputs and outputs to inpatient mental healthcare services as well as key environmental factors employing two different approaches to dealing with environmental variables, first as uncontrolled inputs and then in a two-stage DEA bootstrap model. We test for the sensitivity of the specification of the environmental adjusters in the DEA models by applying various permutations of the environmental variables. The baseline model of inpatient mental healthcare which excludes environmental adjusters appears a slightly more reasonable model with which to examine performance, at least of existing albeit
imperfect indicators, since external factors used in the two-stage model, are not contributing significantly to countries’ ability to deliver technical efficiency.

We find very wide confidence intervals for the country rankings of the 4 DEA models presented. There is a small clutch of poorer performing countries that produce slightly more stable country rankings with narrower confidence intervals. However for the majority of countries, it is impossible to distinguish differences in relative ranking. It is meaningless, therefore, to attempt literal interpretations of country efficiency scores and rankings, particularly for the top performing countries where confidence intervals are wider. The results for the Cluster Analysis show greater complementarity with DEA for those countries grouped at the lower end of the league table where confidence limits were slightly tighter. Cluster Analysis appears to be feasible as a complementary exploratory tool when efficiency rankings are more secure.

Given the variability in country rankings, it would be imprudent for policymakers to focus on the point estimate for their country. Any policy response to this type of analysis should be tempered by the individual country circumstances and action should only be taken after more detailed investigation. DEA and Cluster Analysis are useful diagnostic tools and the methods allow us to indicate where further policy scrutiny may be warranted. These methods may facilitate a dialogue with policymakers about understanding the causes of any anomalies in the data.

International health system comparisons are still at a developmental stage and there is a need to make policymakers aware of the strengths and limitations of the data, methods and results of such comparisons. For instance, some countries in the ‘inefficient’ group may have directed policy effort away from inpatient care, with a focus on people with more severe mental disorders, towards the majority of patients with more mild to moderate mental illness who are treated in ambulatory settings. An analysis which focuses on inpatient care cannot examine these trade-offs.

A further scrutiny of data by policymakers may for instance reveal that indicators are not measured consistently across countries. For instance, in the Netherlands, the measure of beds includes beds in institutions for long term care patients with mental health problems with a low intensity of care. These beds might be included in the social care sector in other countries. Unpicking the data in this way allows policymakers to consider the appropriateness of either improving bed use or improving the measurement of indicators on bed use.

There remains considerable variation in the extent to which national data collection enables comparisons. Improved communication of the data requirements for mental health system performance comparison is a step towards enhancing the utility of such exercises [32].

6. Conclusions

While there are clear limitations to this analysis: only a single output is modelled, no outcomes are considered, only a single year is modelled (when outputs may be the result of cumulative inputs), no adjustment is made for case-mix complexity or quality of care, and there is measurement error in the data, this paper nevertheless has provided a first step towards examining whether DEA with bootstrapping and Cluster Analysis can potentially be used as complementary tools to categorize countries into broad efficiency groupings in this much neglected area. The results suggest the methods may be appropriate for an exploratory analysis to provide a platform for scrutinizing data anomalies and stimulating the search for better data, but given the cross-sectional nature of the data and that they are vulnerable to random fluctuation and measurement error, are not suitable for international comparative analysis.

The Scientific Peer Review Group that commented on the WHR 2000, whilst levelling a number of criticisms on the efficiency analysis, suggested these did not obviate the need for cross-country comparative efficiency studies [28]. One of the essential outcomes of an exercise such as the WHR 2000 [1] was the potential to stimulate the search for improved conceptual models.

A key challenge in comparative analysis of mental health system performance, is the development of a common conceptual framework. The complexity of mental healthcare and the considerable variation between countries in the organization of mental healthcare systems and the infrastructure to deliver care, including the infrastructure available to collect data on indicators of mental health performance, make this a difficult task. Agreeing the priorities for performance assessment is an important first step, since this agenda can only be acted upon if there is agreement on the set of indicators to be incorporated in an efficiency analysis [56].

A key second step is provision of appropriate information on comparative performance in order for policymakers to act upon it. Mental health data has much scope for improvement in terms of quality, consistency and coverage. There is a much greater availability of data in most OECD countries on hospital based care. There is a paucity of data that focuses on recovery-oriented, and patient-centred services in primary and community based settings [13]. Current data at an international level still very much reflects institutional models of care. Evidence suggests investment in mental health information systems is needed to improve measurement of process and outcomes of mental health care, including indicators of effectiveness and safety, data on treatment and procedures, mental morbidity and mortality data [6]. While many OECD countries are investing in information systems to feed this process, much still needs to be done [56]. New data sources such as personal health records will create potential for comparisons of health systems in the treatment of mental disorders, where patient-related factors such as adherence to treatment and lifestyle changes are just as important as proper use of medical interventions and services [32]. The expansion of unique patient identifiers will also mean a radical step forward for many mental health care systems to track patients across settings, across the full care pathway, to assess continuity of care, and quality of prescription and treatment at different levels.

Efforts are being made to improve the level of mental health information in Europe and to support development
of comprehensive mental health information systems for the EU [57]. Such initiatives will support a better future basis for further comparative studies on mental health system performance which can potentially use the methods illustrated in this paper. Comparative analysis of efficiency is still in relatively early stages of development due to its complex, multifaceted and sometimes intangible nature [32] and this is particularly true for mental health care where the additional constraint of poor quality data has led to a virtually non-existent evidence base. While we have been unable to provide robust results of relative mental health system performance, we nevertheless propose an interesting methodological example of how this could potentially be achieved using better data. This paper additionally serves to signal the pressing need for comparative national data to enable meaningful comparisons of the international performance of mental health systems.

Acknowledgements

We would like to thank members of the OECD Mental Health Expert Group for comments on an earlier draft. We are also grateful to two anonymous referees and the editor for helpful comments. The opinions expressed in this article are those of the authors alone; not those of the OECD, nor of its Member countries.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.healthpol.2013.06.011.

References

[1] WHO. The World Health Report 2000: improving health system performance. Geneva; 2001.
[2] Hofmarcher-Holzhaecker M, Winkelmann J, Maas F. EuroREACH Project: improved access to health care data through cross-country comparisons. European Centre for Social Welfare Policy and Research; 2013. http://www.eurocentre.org/detail.php?xml_id=1574
[3] OECD. Health at a glance 2011. Paris; 2011.
[4] Sassi F. Fit not fat. Paris; 2010.
[5] OECD. Improving value for money: measuring quality. Paris; 2010.
[6] Armesto SC, Medeiros H, Wei L. Information availability for measuring and comparing quality of mental health care across OECD countries. In: OECD Health Technical Papers No. 26. Paris; 2008.
[7] STAKES. Establishment of a set of mental health indicators for European Union. Helsinki; 2002.
[8] Hermann RC, Mattek S, Somekh D, Silveyherm H, Goldner E, Glover G, et al. Quality indicators for international benchmarking of mental health care. International Journal for Quality in Health Care 2006;18:31–8.
[9] OECD. OECD health data 2012. Paris; 2012.
[10] WHO. Assessment Instrument for Mental Health Systems (AIMS) Version 2.2. Geneva; 2005.
[11] WHO. Mental health atlas 2011. Geneva; 2011.
[12] Lauriks S, Boster MCA, de Wit MAS, Arah OA, Klazinga NS. Performance indicators for public mental healthcare: a systematic international inventory. BMC Public Health 2012;12:214.
[13] Parmesanwaran S, Sphaeth-Rublee B, Huyyn PF, Pincus HA. Comparison of national mental health quality assessment programs across the globe. Psychiatric Services 2012;63(10):982–8.
[14] Hannigan B, Coffey M. Where the wicked problems are: the case of mental health. Health Policy 2011;101(3):220–7.
[15] Jacobs R, McDaid D. Performance measurement in mental health services. In: Smith PC, Mossialos E, Leatherman S, editors. Performance measurement for health system improvement: experiences, challenges and prospects. Cambridge: Cambridge University Press; 2009. p. 426–71.
[16] WHO. The global burden of disease – 2004 update. Geneva; 2008.
[17] Horton R. Launching a new movement for mental health. The Lancet 2007;370:806.
[18] OECD. Sick on the job: myths and realities about mental health and work. Paris; 2011.
[19] Knapp M, McDaid D, Amaddeo F, Constantopoulos A, Oliveira M, Salvador-Carulla L, et al. Financing mental health care in Europe. Journal of Mental Health 2007;16:167–80.
[20] OECD. Restoring public finances. Paris; 2011.
[21] Jacobs R, Smith P, Street A. Measuring efficiency in health care: analytic techniques and health policy. Cambridge: Cambridge University Press; 2006.
[22] Simar L, Wilson P. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics 2007;136:31–64.
[23] Everitt BS, Landau S, Leese M, Stahl D. Cluster analysis, 5th ed. Chichester: John Wiley & Sons, Ltd.; 2011.
[24] Schinnar AP, Kamis-Gould E, Delucia N, Rothbard AB. Organizational determinants of efficiency and effectiveness in mental health care partial programs. Health Services Research 1990;25(2):387–420.
[25] Ozcan YA, Merwin E, Kwangsou Lee K, Morrissey JP. Benchmarking using DEA: the case of mental health organisations. In: Brandeau ML, Sainfort F, Pierskalla WP, editors. A handbook of methods and applications series: international series in operations research & management science, vol. 70. 2004.
[26] Kontodimopoulos N, Thalia Bellali T, Labiris G, Niakas D. Investigating sources of inefficiency in residential mental health facilities. Journal of Medical Systems 2006;30:169–76.
[27] OECD. Health care systems: efficiency and policy settings. Paris; 2010.
[28] Anand S, Ammar W, Evans T, Hasegawa T, Kissimova-Skarbek K, Langer A, et al. Report of the scientific peer review group on health systems performance assessment. In: Murray CJL, Evans DB, editors. Health systems performance assessment: debates, methods, and empiricism. Geneva: World Health Organisation; 2003. http://www.who.int/health-systems-performance/sgp/hstc14_efficiency.pdf
[29] Lee BL. Efficiency of research performance of Australian universities: a reappraisal using a bootstrap truncated regression approach. Economic Analysis and Policy 2011;41(3):193–203.
[30] Woliszczak-Derlacz J, Parteka A. Efficiency of European public higher education institutions: a two-stage multicountry approach. Sciento- metrics 2011;89:998-917.
[31] Jones AM. Panel data methods and applications to health economics. In: HEDG Working Paper 07/18. University of York; 2007.
[32] Hofmarcher MM, Smith PC, editors. The Health Data Navigator. Your toolkit for comparative performance analysis. A EuroREACH product. Vienna; 2013.
[33] Chisholm D, Sanderson K, Ayuso-Mateos JL, Saxena S. Reducing the global burden of depression: population-level analysis of intervention cost-effectiveness in 14 world regions. British Journal of Psychiatry 2004;184:393–403.
[34] Gutierrez-Recacha P, Chisholm D, Haro JM, Salvador-Carulla L, Ayuso-Mateos JL. Cost-effectiveness of different clinical interventions for reducing the burden of schizophrenia in Spain. Acta Psychiatrica Scandinavica Supplementum 2006:29–38.
[35] Thomas JW, Guire KE, Horvat GG. Is patient length of stay related to quality of care? Hospital & Health Services Administration 1997;42(4):489–507.
[36] OECD. Health at a Glance: Europe 2012. Paris; 2012.
[37] Chilingerian JA, Sherman HD. Health-care applications: from hospital to physicians, from productive efficiency to quality frontiers. In: Cooper WW, Seiford LM, Zhu J, editors. Handbook on data envelopment analysis, 2nd ed. London: Springer; 2011. p. 445–97.
[38] Institute of Medicine. Improving the quality of health care for mental and substance use conditions. Washington, DC: The National Academies Press; 2006.
[39] Taylor R, Page A, Morrell S, Carter G, Harrison J. Socio-economic differentials in mental disorders and suicide attempts in Australia. British Journal of Psychiatry 2004;185:486–93.
[40] Dalgaard OS, Arinstein M, Rognerud M, Johansen R, Zahl PH. Education, sense of mastery and mental health: results from a nationwide health monitoring study in Norway. BMC Psychiatry 2007; 7:20.
[41] Cooper WW, Seiford LM, Tone K. Data envelopment analysis: a comprehensive text with models. In: Applications, references and DEA-solver software. Boston: Kluwer Academic Publishers; 2000.

[42] Smith PC. Model misspecification in data envelopment analysis. Annals of Operations Research 1997;73:233–52.

[43] Orme C, Smith PC. The potential for endogeneity bias in data envelopment analysis. Journal of the Operational Research Society 1996;47:73–83.

[44] Pedraja-Chaparro F, Salinas-Jiménez J, Smith P. On the quality of the data envelopment analysis model. Journal of the Operational Research Society 1999;50:636–44.

[45] Simar L, Wilson P. A general methodology for bootstrapping in nonparametric frontier models. Journal of Applied Statistics 2000;27:779–802.

[46] Olson K, Vu L. Economic efficiency in farm households: trends, explanatory factors, and estimation methods. Agricultural Economics 2009;40:587–99.

[47] Simar L, Wilson P. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. Management Science 1998;44:49–61.

[48] Jensen U. Is it efficient to analyse efficiency rankings? Empirical Economics 2000;25:189–208.

[49] Street A. How much confidence should we place in efficiency estimates? Health Economics 2003;12:895–907.

[50] Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 1984;30:1078–92.

[51] Hollingsworth B, Smith P. The use of ratios in data envelopment analysis. Applied Economics Letters 2003;10(11):733–5.

[52] Ruggiero J. On the measurement of technical efficiency in the public sector. European Journal of Operational Research 1996:90: 533–65.

[53] Ruggiero J. Non-discretionary inputs in data envelopment analysis. European Journal of Operational Research 1998;111: 461–9.

[54] Mooi E, Sarstedt M. A concise guide to market research: the process, data and methods using IBM SPSS statistics. Heidelberg: Springer-Link; 2011.

[55] Wilson P. FEAR 1.15 user’s guide. South Carolina: Clemson University; 2010.

[56] Hussey PS, de Vries H, Romley J, Wang MC, Chen SS, Shekelle PG, et al. A systematic review of health care efficiency measures. Health Services Research 2009;44(3):784–805.

[57] Lavikainen J, Fryers T, Lehtinen V, STAKES and European Union. Improving mental health information in Europe: Proposal of the MINDFUL project. Helsinki, Finland: STAKES, EU and MINDFUL; 2006. http://www.stakes.fi/pdf/mentalhealth/Mindful_verkkoversio.pdf