Epidemiology and Psychiatric Sciences (2015), **24**, 78–89.
doi:10.1017/S204579601300067X

ORIGINAL ARTICLE

Cost drivers of inpatient mental health care: a systematic review

J. Wolff[1,2]*, P. McCrone[1], L. Koeser[1], C. Normann[2] and A. Patel[1]

[1] *King's College London, Institute of Psychiatry, Health Service and Population Research, CEMPH, London, UK*
[2] *Department of Psychiatry and Psychotherapy, University Medical Centre Freiburg, Freiburg im Breisgau, Germany*

Aims. New reimbursement schemes for inpatient mental health care are imminent in the UK and Germany. The shared intention is to reflect cost differences between patients in reimbursement rates. This requires understanding of patient characteristics that influence hospital resource use. The aim of this review was to show which associations between mental health care per diem hospital costs and patient characteristics are supported by current evidence.

Methods. A systematic review of the literature published between 1980 and 2012 was carried out. The search strategy included electronic databases and hand-searching. Furthermore, reference lists, citing articles and related publications were screened and experts were contacted.

Results. The search found eight studies. Dispersion in per diem costs was moderate, as was the ability to explain it with patient characteristics. Six patient characteristics were identified as the most relevant variables. These were (1) age, (2) major diagnostic group, (3) risk, (4) legal problems, (5) the ability to perform activities of daily living and (6) presence of psychotic or affective symptoms. Two non-patient-related factors were identified. These were (1) day of stay and (2) treatment site.

Conclusions. Idiosyncrasies of mental health care complicated the prediction of per diem hospital costs. More research is required in European settings since transferability of results is unlikely.

Received 5 September 2013; Revised 18 October 2013; Accepted 22 October 2013; First published online 16 December 2013

Key words: Cost analysis, hospitals, mental disorders, prospective payment system.

Introduction

Inappropriate hospital reimbursement can lead to inefficient delivery of care. Hospitals may reduce costs at the expense of quality, cut length of stay (LOS) for pecuniary reasons and avoid treatment of high-cost patients (Rupp *et al.* 1984; Jencks *et al.* 1987; Essock & Norquist, 1988). Such responses can be stronger in mental health than in medical and surgical care (Lave & Frank, 1990). Reimbursement schemes are currently being developed in the UK and in Germany that aim to address such inadequate incentives by adjusting rates to differences in resource use between classes of patients (case-mix). These classes should be relatively homogenous in terms of costs in order to set an adequate rate of reimbursement for all patients within a class (McCrone, 1994). Therefore, it is necessary to identify patient characteristics that influence resource use.

Inpatient mental health care frequently requires longer admissions than somatic health care and LOS is a relevant determinant of costs per stay (Suarez *et al.* 2011; Tulloch *et al.* 2011). Research surrounding the unsuccessful implementation of Diagnosis-related Groups for mental health care during the 1980s in the USA found that it was difficult to predict LOS from patient characteristics (McCrone & Phelan, 1994). The attention has now shifted more towards per diem classification systems, i.e. reimbursement for each day of stay instead of a lump sum for the whole episode, and such systems are used in the USA and in the process of development and implementation in Germany.

The main aim of this review was to show which associations between mental health care per diem hospital costs and patient characteristics are supported by current evidence. A secondary aim was to identify other, non-patient-related factors that have implications on costs. These aspects are not only relevant for policy makers and concerned stakeholders, but also inform decision-making in routine clinical care, for instance, by substantiating staff planning and performance benchmarks, and evaluating service portfolios.

* Address for correspondence: J. Wolff, Department of Psychiatry and Psychotherapy, University Medical Centre Freiburg, Hugstetter Straße 49, 79106 Freiburg im Breisgau, Germany.
(Email: jan.wolff@kcl.ac.uk)

Methods

A systematic review of the literature published between 1980 and 2012 was carried out. Relevant studies analysed psychiatric, psychotherapeutic or psychosomatic inpatient care for adult patients with mental or behavioural disorders. The outcome of interest was per diem costs in relation to patient characteristics. Studies were excluded if the scope and detail of measurement were not sufficient, i.e. improvements on existing processes of unit cost calculation based on routine data in the UK (Payment by Results, 2012) and Germany (InEK, 2010) for adjusting hospital reimbursement. A sufficient scope of measurement required consideration of patient-specific use of staff time including routine care because this was estimated to be the main cost factor in inpatient mental health care (Cotterill & Thomas, 2004; Moreno, 2007). A sufficiently detailed measurement required a micro-costing approach that employed physical units of resource use, such as minutes of time for personnel costs, instead of aggregating or using proxy measures, such as bed-days or acuity scores. There were neither restrictions on formal design, such as randomised clinical trials, nor on language or type of publication.

Search strategy

The search strategy included electronic databases and hand-searching. Reference lists, citing articles and related articles were searched in a second step. MEDLINE, EMBASE, PSYCINFO and HMIC were accessed via OvidSP. CINAHL, PSYNDEX and ECONLIT were accessed via EBSCOhost. CENTRAL, NHS-EED and the Health Technology Assessment Database (HTA) were accessed via Cochrane Library. A search query was initially developed for MEDLINE and subsequently translated to other databases. The full search strategy is provided in the online appendix. The database searches were conducted between 25th and 30th of August 2012 and current awareness alerts implemented until 1st January 2013. Hand-searching included all issues between 2002 and 2012 of *Acta Psychiatrica Scandinavica*, *British Journal of Psychiatry*, *Epidemiology and Psychiatric Sciences*, *European Journal of Health Economics*, *Journal of Mental Health Policy and Economics*, and the German hospital management journals *Führen und Wirtschaften*, *Das Krankenhaus* and *Krankenhausumschau*. Google Scholar was used to identify citing records and to screen the first 20 offered 'related citations' of every record identified as relevant beforehand. Finally, a list of identified relevant records was sent to experts and authors in the field of interest to ask for unpublished or on-going research.

Study selection and data extraction

After clearing duplicate listings of the same record, titles and abstracts were screened by a single researcher (JW) to exclude records that were clearly irrelevant to the research question in stages 1 and 2, respectively. A second researcher (LK) reassessed a 10% random sample of excluded records from each stage in order to reassure the quality of the screening process without finding inappropriate decisions. Both researchers were health economists. The reassessment was conducted blindly, i.e. without knowing the name of the journal, the authors or their affiliations. Full texts were retrieved for the remaining records and two researchers independently and blindly decided eligibility. Disagreements were resolved by discussion. Two authors of studies had to be contacted in order to clarify fulfilment of inclusion criteria.

Multiple publications of the same study were merged before data extraction. The small number of relevant publications allowed an in-depth discussion of data. One author was contacted in order to clarify applied methods. Non-English/German language publications were translated before extraction of data. If necessary, reported results were recalculated to address the research question.

Results

The literature search yielded eight studies reported in 17 publications. The selection process is illustrated in Fig. 1. Precision of the database search was comparatively low. Eight publications were eventually included from 15 536 unique records identified in the database search (Autio, 1987; Fries *et al.* 1990, 1993; Cromwell *et al.* 2004, 2005*b*; Iglesias & Alonso Villa, 2005; Cromwell & Maier, 2006; Drozd *et al.* 2006). Eight of 17 publications were grey literatures (Yamauchi, 1997*a*, *b*; Buckingham *et al.* 1998*a*, *b*; Hirdes *et al.* 2002; Gaines *et al.* 2003; Cromwell *et al.* 2005*a*; Daniel, 2008). One publication (Eagar *et al.* 2004) was available in databases but not identified by the search query because it was not indexed with key words or pertinent free texts and missed the search terms and subject headings used for the concept 'inpatient setting'.

Study characteristics

Basic features of study design are provided in Table 1. Six studies were conducted in multiple sites with large sample sizes (Group A) and two studies at a single site with smaller sample sizes (Group B). All studies in Group A were reported in multiple publications (Study 1: Cromwell *et al.* 2004, 2005*a*, *b*; Cromwell &

https://doi.org/10.1017/S204579601300067X Published online by Cambridge University Press

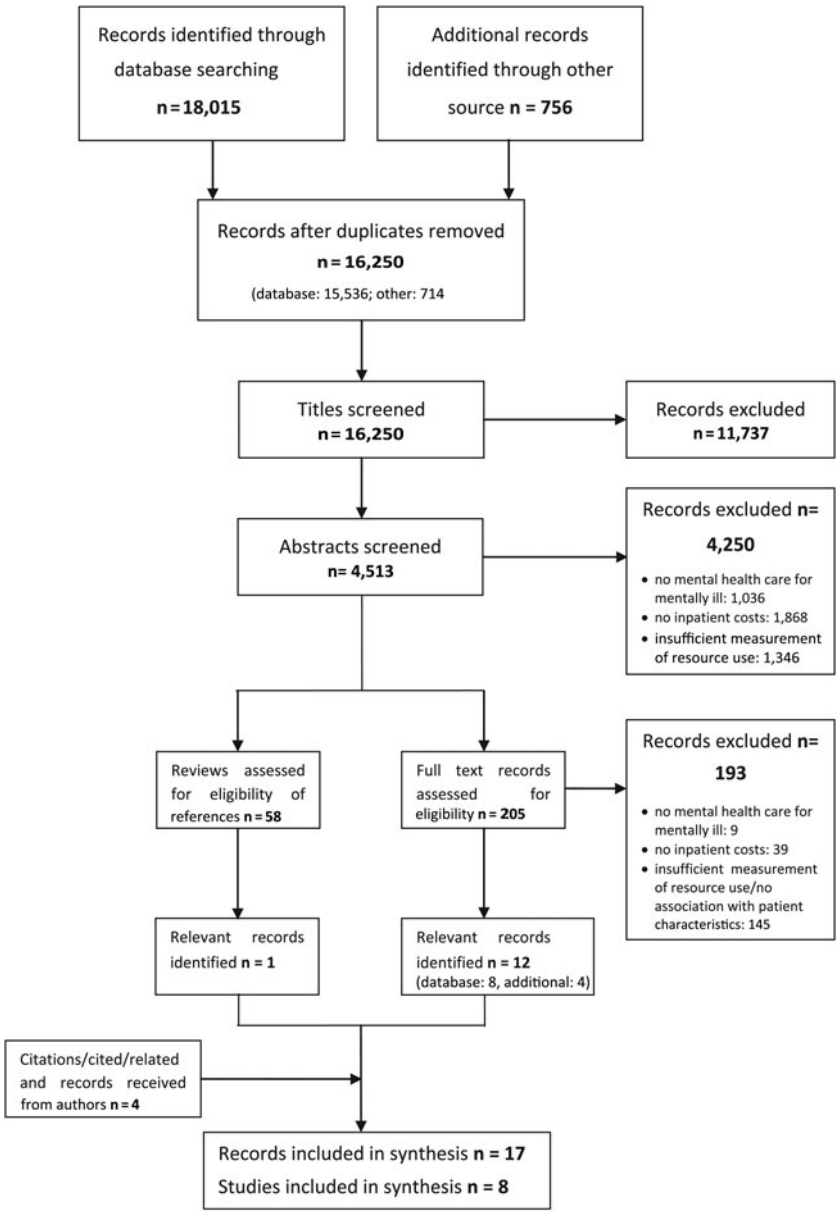

Fig. 1. Flow chart literature search.

Maier, 2006; Drozd *et al.* 2006; Study 2: Gaines *et al.* 2003; Eagar *et al.* 2004; Study 3: Hirdes *et al.* 2002; Daniel, 2008; Study 4: Buckingham *et al.* 1998*a*, *b*; Study 5: Yamauchi, 1997*a*, *b*; Study 6: Fries *et al.* 1990, 1993). A total of 15 613 patient episodes were included with a range of 42–6554. All studies included personnel costs and measured patient-specific staff time including routine care, since this was a required inclusion criterion. Cromwell *et al.* (2005*a*), Gaines *et al.* (2003) and Buckingham *et al.* (1998*b*) included additional indirect and overhead costs and services from ancillary cost centres (e.g. radiology, pharmacy and pathology). Resource use measurement differed with regard to methods and duration of measurement. Continuous self-reports by staff to log patient-specific work time were the most prevalent approach. Cromwell *et al.* (2005*a*) complemented self-reports with data from patient observations for allocation of time among patients. Gaines *et al.* (2003) did not continuously note activities but summarised time per patient at the end of each shift. Autio (1987) noted patient activities at randomised 10-min intervals and extrapolated distribution of staff time. The duration of measurement ranged from 1 day to 6 months.

For resource use valuation, average wages per professional group were used or staff time served as the basis for top-down allocation of personnel costs. Cromwell *et al.* (2005*a*) weighted staff time by average

https://doi.org/10.1017/S204579601300067X Published online by Cambridge University Press

Table 1. *Study design*

		Year of data	Location	Treatment sites	Patient episodes	Patients	Costs included	Measurement of staff time	Valuation of resource use
GROUP A									
1	Cromwell *et al.*	2001–2003	USA	40	834	Medicare	Direct care staff, ancillary indirect/overhead	Self-reported logs, direct observation, 7 days	CMI multiplied with average costs
2	Gaines *et al.*	2003	New Zealand	8	2848	No restriction	medical/health staff, ancillary, indirect/ overhead	Self-reported summaries (nurses), 6 month	Top-down allocation
3	Hirdes *et al.*	1999–2000	Canada	34	1998	No restriction	Clinical staff excl. physicians	Self-reported logs, 1 day/7 days	Avg. wages
4	Buckingham *et al.*	1996	Australia	22	6554	No restriction	Direct care staff, indirect/ overhead	Self-reported logs, 3 months	Top-down allocation
5	Yamauchi	1993	Japan	17	2284	No restriction	Physicians, nurses, therapists	Self-reported logs, 1 day/7 days	Avg. wages
6	Fries *et al.*	Unclear	USA	51	890	Long-staying	Nurses, physicians, psychologists, other therapists	Self-reported logs, 1 day/7 days	Avg. wages
GROUP B									
7	Iglesias & Alonso Villa	2003	Spain	1	163	Long-staying	Nurses, aides	Self-reported logs, 7 days	None
8	Autio	Unclear	USA	1	42	No restriction	Nurses, counselors	Obs. patient sampling, rand. 10 min interval, ten shifts	None

Ancillary = e.g. radiology, pharmacy, pathology; CMI = case-mix index; obs.: = observational; rand. = randomised.

wages per professional group and calculated an index by putting each patient in relation to the average. This index was multiplied by the facility-wide routine per diem costs, which included cost types not necessarily related to staff time. Iglesias & Alonso Villa (2005) and Autio (1987) did not value resource use monetarily.

All studies in Group A used regression trees for the analysis of data. This non-parametric method recursively partitions the sample into smaller subgroups by identifying variables that best classify patients into subgroups of homogenous per diem costs (Breiman *et al.* 1984).

Cost structure

Table 2 provides information considering cost structures of each study, required to understand the implications of methodological differences on results. Cromwell *et al.* (2005*a*), Gaines *et al.* (2003) and Buckingham *et al.* (1998*b*) distinguished between variable resource uses, e.g. staff time, and resource use that was constant per patient day, e.g. overhead costs. Other studies did not allocate constant per diem costs but either excluded non-variable resource use (Autio, 1987; Yamauchi, 1997*a*; Hirdes *et al.* 2002; Iglesias & Alonso Villa, 2005) or allocated fixed/similar resource use on a variable basis (Fries *et al.* 1990). Studies with fixed shares of daily resource use found less difference in costs between patients than studies that did not allocate constant resource use. Measures of dispersion, i.e. the coefficient of variation and the interquartile range, increased monotonically with an increasing share of variable costs. In order to allow quantitative comparison of results, an adjustment factor was calculated for each study to rescale the results, namely the reciprocal of variable costs shares. These were used to approximate results in later comparative analysis[†]. The number of patient classes created with regression trees ranged from 4 to 47. Studies used between 2 and 45 variables for classification and explained variance ranged from 10 to 40%.

Patient characteristics associated with resource use

Of 49 unique patient variables used for classification in Group A, six were selected as most relevant, because they were used by at least two studies within the first three levels of their regression trees. These were the ability to perform activities of daily living (ADL), age, legal problems, major diagnostic group (MDG), experiencing psychotic or affective symptoms and factors related to aggression or danger against self or others (risk). Their univariate explanatory power in terms of variance of per diem costs is shown in Table 3. Effect sizes measured by the ratio of the constant prediction functions for the most to the least costly group that were differentiated by these variables are provided in Table 4. In addition, rescaled ratios are given adjusted to differences in shares of fixed per diem costs. In split-level 1, main effects are reported. Afterwards, figures in split levels 2 and 3 are reported and these represent interaction effects between the bold printed variable and the variable in brackets.

Age

Age was a poor predictor in the linear regression models with a maximum of 3.3% explained variance (Table 3). However, it had the strongest cost split in Cromwell *et al.* (2005*a*) and Buckingham *et al.* (1998*b*), who placed cut-off points at the specific age thresholds (Table 4). Cromwell *et al.* (2005*a*) separated patients that were at least 65 years of age from others. The rescaled ratio was 1.37, meaning that older patients were 37% more costly than others. Buckingham *et al.* (1998*b*) split the complete sample into three age groups. Patients of at least 65 years of age were 4% more costly than patients between 34 and 64 years of age. However, adult patients of less than 34 years of age formed the most expensive group with a rescaled ratio of 1.52 compared with patients between 34 and 64 years of age.

Major diagnostic group

MDG accounted for the second highest amount of variance in the linear regression models (6.4%) after ADL (6.9%). Cromwell *et al.* (2005*a*) and Hirdes *et al.* (2002) used MDG as forced splits in regression trees for the complete sample and separated patients into five and seven subgroups, respectively. A forced split means that this variable was not automatically detected on statistical grounds but chosen by the authors for other reasons, such as clinically meaningful classifications. Furthermore, Buckingham *et al.* (1998*b*) statistically detected MDG as having the second best cost split after age. The case-mix index for MDG within their parent groups is provided in Fig. 2. Dementia and mood disorders patients were consistently more costly than average as were schizophrenia and substance-related disorder patients less costly than average. Hirdes *et al.* (2002) found patients with organic disorder to have higher costs than average. However, Buckingham *et al.* (1998*b*) found only organic disorder

[†] Adjustment factor in Cromwell *et al.*: 1/0,71 = 1.41; published difference in per diem costs = 26%; rescaled difference = 1.41*26% = 37%

https://doi.org/10.1017/S204579601300067X Published online by Cambridge University Press

Table 2. *Cost structure of included studies*

		Variable costs (%)	Fixed costs (%)	CV	IQR	Class. groups	Variables used for class.	R^2
GROUP A								
1	Cromwell *et al.*	71	29	0.53	2.04	13–25	45	24–40%
2	Gaines *et al.*	18	82	0.31	1.34	4	2	10%
3	Hirdes *et al.*	100	0	0.83	na	47	22	26%
4	Buckingham *et al.*	29	71	0.35	1.51	12	5	19%
5	Yamauchi	100	0	0.90	na	18	9	27%
6	Fries *et al.*	100	0	0.59	na	6	6	11%
GROUP B								
7	Iglesias & Alonso Villa	100	0	0.82	na	na	na	na
8	Autio	100	0	0.66–1.65	na	na	na	na

CV = coefficient of variation (ratio of standard deviation to mean); IQR = interquartile range (ratio of third quartile to first quartile); class. = classification; R^2 = explained variance.

patients who were at least 65 years of age to have higher costs than average. Other patients with organic disorders were less expensive than average.

Risk and legal problems

Risk and legal problems were both identified by one study as resulting in the strongest cost split for the complete sample. They reached a maximum explained variance in the linear regression models of 5.7 and 1.5%, respectively. Moreover, patients with legal problems were more than three times as expensive as other patients in Gaines *et al.* (2003), who did not report results of univariate regression.

Results in Group B

Results in Group B are reported separately because different analytical methods were used. Iglesias & Alonso Villa (2005) grouped patients according to a previously established classification system (Fries *et al.* 1989) and compared differences in mean resource use. They found a positive association between the degree of dependency in ADL and resource use. In patients with severe behavioural problems, the group with any dependency were 49% more resource intensive than patients without dependency. In patients with reduced physical functioning, the group with the highest dependency was more than twice as resource intensive as patients without dependency. Autio (1987) analysed six patient variables, namely nurses' subjective estimate of a patient's resource intensity, legal status, social support, past psychiatric hospitalisation, secondary medical condition and level of adaptive functioning. A stepwise regression approach identified only social support to be significantly predictive of staff time use accounting for 11% of its variance.

Other factors influencing resource use

Two associations between non-patient-related factors and per diem costs were identified by more than one study. These were day of stay and treatment site.

Table 3. *Explained variance of individual variables*

	ADL	Age	Legal problems	MDG	Psychot./affect. sym.	Risk
2 Gaines *et al.*	–	–	–	6.4%	–	–
3 Hirdes *et al.*	–	–	–	2.4%	–	–
4 Buckingham *et al.*	–	2.0%	1.5%	5.7%	–	1.7–2.1%
5 Yamauchi	6.9%	3.3%	–	2.3%	–	4.3%
6 Fries *et al.*	1.4%	1.1%	–	–	3.0%	3.9–5.7%

ADL = Activities of Daily Living; MDG = major diagnostic group; Psychot./affect. sym. = psychotic or affective symptoms.

https://doi.org/10.1017/S204579601300067X Published online by Cambridge University Press

Table 4. *Effect size measured by ratio of high-to-low cost patient groups*

			Effect size	Rescaled
L1	Age	Cromwell *et al.* (complete sample)	1.26	1.37
		Buckingham *et al.* (complete sample)	1.15	1.52
	Major diagnostic group*	Cromwell *et al.* (complete sample)	1.31	1.44
		Hirdes *et al.* (complete sample)	1.76	1.76
	Risk	Fries *et al.* (complete sample)	1.60	1.60
	Legal problems	Gaines *et al.* (complete sample)	1.39	3.16
L2	ADL	Cromwell *et al.* (dementia)	1.15	1.21
		Cromwell *et al.* (schiz.)	1.42	1.60
		Hirdes *et al.* (organic disorders)	1.31	1.31
		Yamauchi (2nd month–6th month)	2.24	2.24
		Yamauchi (after 6th month)	2.77	2.77
	Age	Cromwell *et al.* (mood disorders)	1.20	1.28
	Psychot./affect. sym.	Fries *et al.* (no risk)	1.16	1.16
	Major diagnostic group	Cromwell *et al.* (age <65)	1.18	1.25
		Buckingham *et al.* (age >64)	1.14	1.50
		Buckingham *et al.* (age 34–64)	1.18	1.63
		Buckingham *et al.* (age <34)	1.22	1.77
	Risk	Cromwell *et al.* (subs. rel)	1.26	1.37
		Hirdes *et al.* (other disorders)	1.52	1.52
L3	ADL	Cromwell *et al.* (age >64 & no detox)	1.13	1.18
		Buckingham *et al.* (age 65+ & organic dis.)	1.16	1.55
		Buckingham *et al.* (age 65+ & schiz./subs.rel./ret.)	1.18	1.62
		Fries *et al.* (no risk & none psychotic)	1.26	1.26
	Age	Cromwell *et al.* (schiz./high ADL need)	1.14	1.20
		Cromwell *et al.* (subs.rel. & risk)	1.15	1.20
		Cromwell *et al.* (schiz. & low ADL need)	1.16	1.23
	Legal problems	Cromwell *et al.* (age <65 & schiz./subs.rel.)	2.33	2.47
		Cromwell *et al.* (schiz. & low ADL need)	1.19	1.27
		Hirdes *et al.* (others disorders & no risk)	1.02	1.02
		Buckingham *et al.* (age <34 & schiz./organic dis.)	1.47	2.63
	Psychot./affect. sym.	Hirdes *et al.* (mood dis. & day 0–4)	1.37	1.37
		Hirdes *et al.* (mood dis. & day 5+)	1.51	1.51
	Major diagnostic group	Cromwell *et al.* (age >64 & on detox)	1.21	1.30
	Risk	Hirdes *et al.* (others disorders & risk)	1.43	1.43
		Hirdes *et al.* (schiz./psychotic & day 5–730)	1.61	1.61

Continued

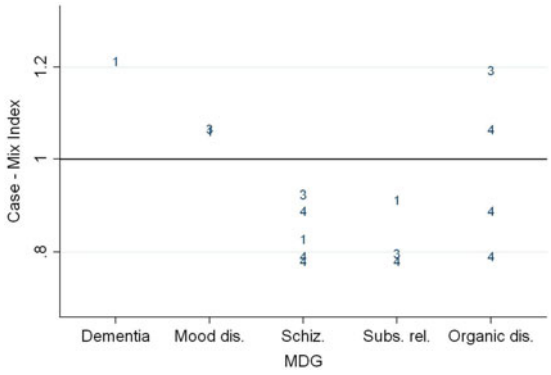

Fig. 2. Case-mix index of major diagnostic groups. Plotted marks: study numbers; Dis. = disorders; Schiz. = schizophrenia; Subs. rel. = substance-related disorders.

Day of stay

Cromwell *et al.* (2005*a*), Hirdes *et al.* (2002) and Fries *et al.* (1993) analysed differences in per diem costs during a patient's admission. Cromwell *et al.* (2005*a*) found that days 1 and 2 were 30 and 14% more cost intensive (rescaled) than average, respectively. Days 3 and 4 were 4% more costly. After day 4, the costs were below average. Days 5 to 7 accounted for 92% of average per diem costs and days 8 to 14 were at 91%. After day 14, per diem costs were 11% below average.

Hirdes *et al.* (2002) found schizophrenia and other psychotic disorder patients to incur 76% higher per diem costs than average during their first 5 days of stay. Daily costs between days 5 and 730 were below average at 98%. After the first 2 years, per diem costs were at 78% of average costs. A comparable decline was found for mood disorder patients, where per diem costs were 50% higher than average costs during the first 5 days of stay and 12% lower than average costs afterwards.

Fries *et al.* (1993) found the greatest decline of per diem costs in personality disorder patients, which accounted for only 33% of initial daily costs after 1 week. Organic disorder patients incurred 18% higher per diem costs during the first 3 days than in the following 11 days. Other patients incurred a decline of <8% after the first week.

Treatment site

Cromwell *et al.* (2005*a*), Buckingham *et al.* (1998*b*) and Fries *et al.* (1990) carried out linear regressions of costs on treatment site. They found strong associations, accounting for 28, 25 and 39% of variance in per diem costs, respectively. In comparison, the strongest patient variables in Buckingham *et al.* (1998*b*) and

Table 4. *Continued*

	Effect size	Rescaled
Hirdes *et al.* (schiz./psychotic & day 730+)	1.86	1.86
Buckingham *et al.* (age 34–64 & schiz./organic dis.)	1.08	1.28
Yamauchi (1st month & no/moderate hallucin.)	2.73	2.73
Yamauchi (1st month & severe/ extrem. hallucin.)	2.02	2.02
Yamauchi (after 6th month & low ADL)	1.81	1.81

Regression trees do not provide confidence intervals or probability levels; * = Forced split; L1 = Split level 1; Activities of Daily Living; Psychot./affect. sym. = psychotic or affective symptoms; schiz = schizophrenia; subs.rel. = substance-related disorders; dis. = disorder; ret. = mental retardation; hallucin. = hallucinations; other disorders = other than schiz./psychotic, organic, mood, personality, eating and substance-related disorder.

https://doi.org/10.1017/S204579601300067X Published online by Cambridge University Press

Fries *et al.* (1990), namely MDG and risk, accounted for only 5.7% of variance in per diem costs.

Breaking it down to treatment site's properties, Cromwell *et al.* (2005*a*) found facility size, measured by the average daily patient census, and teaching intensity, measured by the ratio of interns to the average daily patient census, to be significantly related to per diem costs after controlling for patient variables and the day of stay. Regression coefficients indicated a decrease of 1.5% in per diem costs per 10% increase of facility size. Regression coefficients for teaching intensity indicated an 11% increase in resource intensity for a 0.1 increase of interns per patient (time of interns was excluded from resource use).

Discussion

The main aim of this review was to show which associations between per diem hospital costs of mental health care and patient characteristics are supported by current evidence. A secondary aim was to identify other, non-patient-related factors that have implications on costs. The dispersion in per diem costs was moderate, as was the ability to explain it with patient characteristics. Six patient characteristics were identified as the most potent variables with regard to explanation of cost dispersion. These were age, MDG, risk, legal problems, ADL and presence of psychotic or affective symptoms. Furthermore, two non-patient-related factors were identified. These were day of stay and treatment site.

Strength and weaknesses of the study

A systematic search identified eight studies reported in 17 publications, evenly distributed between papers in academic journals and grey literature. The search strategy was comparatively comprehensive with regard to the number of sources and throughput of records. The precision of the database search was low due to the nature of the research question and methodological requirements for inclusion of studies. No language restrictions were applied. This allowed inclusion of a rarely cited but relevant study in Japanese. The amount of patient variables under investigation necessitated selection of the most relevant findings. Since there is nothing like a gold-standard approach for this decision, new criteria had to be defined.

Despite methodological differences, the quality of studies in Group A generally fulfilled agreed standards (Drummond & Jefferson, 1996; Evers *et al.* 2005). However, three aspects should be mentioned with regard to resource use measurement, valuation and analysis. First, not all studies included all resource use associated with inpatient treatment. In particular, exclusion of specific staff groups, such as physicians and psychologists, from resource use measurement in Gaines *et al.* (2003) and Hirdes *et al.* (2002) must be kept in mind when interpreting results because their distribution of work time among patients might be structurally different to other staff groups. Moreover, only Cromwell *et al.* (2005*a*), Gaines *et al.* (2003) and Buckingham *et al.* (1998*b*) included resource use beyond staff time. Although personnel costs represented the largest cost share, consideration of other costs, such as pharmaceuticals or diagnostic procedures, could have revealed interesting differences between groups of patients or between the first and the following days of stay.

Second, differences in valuation of resource use hampered comparison of results. Up to fivefold differences in variable cost shares affected results as shown by measures of cost dispersion and approximation was required in order to reasonably compare results. There were insufficient data provided to indisputably rescale results. Therefore, a heuristic narrowing approach was taken by using the reciprocal of the mean variable cost share as adjustment factors.

Third, comparative synthesis of results was possible because most studies used regression trees. In contrast to other methods, which make it difficult to incorporate high-order statistical interactions, regression trees are inherently interactive. This allowed reporting of differential effects for the complete sample and for each subgroup of patients. However, the widely used method of regression trees does not provide confidence intervals or probability levels. Instead, the generalisability of model outputs to other datasets was addressed with measures specific to the different tree algorithms, such as pruning of unreliable branches of the tree and cross-validation (Jensen & Cohen, 2000). It is also important to mention that the studies did not group solely on statistical grounds but included others aspects in order to create classes usable for reimbursement purposes, such as clinical meaningfulness, avoidance of gaming and reduction of administrative burden. It was not possible to take all these decisions into account since they were not consistently reported. However, focusing on classes that are usable avoids meaningless conclusions, such as finding similar costs between very different patients, e.g. dementia in old men and anorexia in young women (Buckingham *et al.* 1998*b*).

Studies in Group B did not monetarily value resource use. In addition to this limitation, a problem in Iglesias & Alonso Villa (2005) was lack of sufficient significance testing limiting the degree of certainty in its results. The main limitation in Autio (1987) was the small sample size, which limited significant findings to a single variable.

https://doi.org/10.1017/S204579601300067X Published online by Cambridge University Press

Results in relation to prior research

Classification systems showed a comparatively low performance in explaining differences in per diem costs with patient characteristics. The best performing model in Cromwell *et al.* (2005*a*) explained 40% of variance. However, their model was not a pure patient classification as it included eight facility characteristics plus services variables, such as electroconvulsive therapies. In comparison, Fries *et al.* (1994) also carried out a study in 228 nursing home units and measured patient-specific per diem staff time of 7658 residents during 1 day for regular and 1 week for seldom used activities. Their classification system excluded facility characteristics and explained 56% of variance with 44 groups. A possible explanation for lower performance of classifications in mental health is a lack of standardisation of care, meaning that resources allocated to patients could depend less on needs and more on treatment style of caregivers (Weinmann *et al.* 2007; Barbui & Tansella, 2012). This explanation would also be epistemologically coherent with the strong association between treatment site and per diem costs.

A strong influence of provider factors might have reduced the impact of patient characteristics. Indeed, the individual performances of patient characteristics were moderate and comparable with prior research in modelling LOS (McCrone, 1995). However, there was a rather consistent direction of effects in MDG. Generally, schizophrenia and substance-related disorders were less costly than average, whereas dementia and mood disorders were more costly than average. It is also likely, although not reported, that performance of variables interacting in regression trees was superior to global regression models for the complete sample. However, increases have not been very substantial, as indicated by dividing total explained variance by the number of included variables.

Implications and further research

Per diem costs were quite homogenous even before classification. All coefficients of variation were below one. This raises the question as to whether constant per diem rates would be sufficient for reimbursement purpose and preferable with regard to administrative burden.

However, there are at least two reasons why reimbursement should follow an estimate of true costs. First, although dispersion in the complete sample was low, some patients might induce substantially higher costs than others and equal reimbursement would lead to disincentives in the delivery of care. Second, included studies focused on specific resource use instead of covering all relevant aspects of care. This might have disguised true cost differences.

Apparently, classification of patients according to costs in mental health care requires more in-depth research than in other medical disciplines. For instances, differences in surgical costs between appendectomies and heart transplantations are obvious and rather easy to measure. In contrast, mental health care has more complex diagnostic concepts and less agreement on treatment strategies (Lave, 2003), making it more difficult to achieve clear results.

The currently most relevant patient characteristics for cost models were identified by this review. Further research should focus on aspects not included in this review, such as acuteness of disease, which could have been one aspect responsible for declining cost trajectories during admissions. Furthermore, greater delineation by severity class appears promising, since mild episodes should be less resource intensive than very severe episodes.

More research is required in European settings, because it is unlikely that results are transferable across health care systems (McCrone *et al.* 2001; Wahlbeck, 2011). In particular, new reimbursement schemes are currently being introduced in Germany and the UK and evidence will greatly inform such developments.

Acknowledgements

None.

Financial Support

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflict of Interest

None.

Ethical Standard

Not applicable.

References

Autio S (1987). The relationship between patient characteristics and inpatient psychiatric nurse resource utilization. *NLN Publications* **20**, 319–330.

Barbui C, Tansella M (2012). Guideline implementation in mental health: current status and future goals. *Epidemiology and Psychiatric Sciences* **21**, 227–229.

https://doi.org/10.1017/S204579601300067X Published online by Cambridge University Press

Breiman L, Friedman J, Stone CJ, Olshen RA (1984). *Classification and Regression Trees*. Chapman and Hall: Boca Raton.

Buckingham B, Burgess P, Solomon S (1998*a*). *Developing a Casemix Classification for Mental Health Services – Summary*. Commonwealth Department of Health & Family Services: Canberra.

Buckingham B, Burgess P, Solomon S, Pirkis J, Eager K (1998*b*). *Developing a Casemix Classification for Mental Health Services: Volume 1 Main Report*. Commonwealth Department of Health & Family Services: Canberra.

Cotterill PG, Thomas FG (2004). Prospective payment for Medicare inpatient psychiatric care: assessing the alternatives. *Health Care Financing Review* **26**, 85–101.

Cromwell J, Maier J (2006). Economic grand rounds: variation in staffing and activities in psychiatric inpatient units. *Psychiatric Services* **57**, 772–774.

Cromwell J, Maier J, Gage B, Drozd E, Osber D, Richter E, Greenwald L, Goldman H (2004). Characteristics of high staff intensive Medicare psychiatric inpatients. *Health Care Financing Review* **26**, 103–117.

Cromwell J, Gage B, Drozd E, Maier J, Osber D, Evensen C, Richter E, Goldman H (2005*a*). *Psychiatric Inpatient Routine Cost Analysis – Final Report*. RTI International: Waltham.

Cromwell J, Drozd E, Gage B, Maier J, Richter E, Goldman H (2005*b*). Variation in patient routine costliness in U.S. psychiatric facilities. *Journal of Mental Health Policy and Economics* **8**, 15–28.

Daniel I (2008). *Grouper and Weighting Methodology for Adult Inpatient Mental Health Care in Ontario- Summary Report From The JPPC Mental Health Technical Working Group*. Ontario Joint Policy and Planning Committee: Ontario.

Drozd E, Cromwell J, Gage B, Maier J, Greenwald L, Goldman H (2006). Patient casemix classification for medicare psychiatric prospective payment. *American Journal of Psychiatry* **163**, 724–732.

Drummond MF, Jefferson TO (1996). Guidelines for authors and peer reviewers of economic submissions to the BMJ. *British Medical Journal* **313**, 275–283.

Eagar K, Gaines P, Burgess P, Green J, Bower A, Buckingham B, Mellsop G (2004). Developing a New Zealand casemix classification for mental health services. *World Psychiatry* **3**, 172–177.

Essock SM, Norquist GS (1988). Toward a fairer prospective payment system. *Archives of General Psychiatry* **45**, 1041–1044.

Evers S, Goossens M, de Vet H, van Tulder M, Ament A (2005). Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. *International Journal of Technology Assessment in Health Care* **21**, 240–245.

Fries BE, Schneider DP, Foley WJ, Dowling M (1989). Case-mix classification of Medicare residents in skilled nursing facilities: Resource Utilization Groups (RUG-T18). *Medical Care* **27**, 843–858.

Fries BE, Nerenz DR, Falcon SP, Ashcraft ML, Lee CZ (1990). A classification system for long-staying psychiatric patients. *Medical Care* **28**, 311–323.

Fries BE, Durance PW, Nerenz DR, Ashcraft ML (1993). A comprehensive payment model for short- and long-stay psychiatric patients. *Health Care Financing Review* **15**, 31–50.

Fries BE, Schneider DP, Foley WJ, Gavazzi M, Burke R, Cornelius E (1994). Refining a case-mix measure for nursing homes: Resource Utilization Groups (RUG-III). *Medical Care* **32**, 668–685.

Gaines P, Bower A, Buckingham B, Eager K, Green J (2003). *New Zealand Health Classification and Outcomes Study: Final Report*. Health Research Council of New Zealand: Auckland.

Hirdes JP, Fries BE, Botz C, Ensley C, Marhaba M, Pérez E (2002). *The System for Classification of In-Patient Psychiatry (SCIPP): a New Case-Mix Methodology for Mental Health*. InterRAI: Waterloo.

Iglesias C, Alonso Villa MJ (2005). A system of patient classification in long-term psychiatric inpatients: Resource Utilization Groups T-18 (RUG T-18). *Journal of Psychiatric and Mental Health Nursing* **12**, 33–37.

InEK (2010). *Kalkulation von Behandlungsfallkosten- Handbuch zur Kalkulation psychiatrischer und psychosomatischer Leistungen in Einrichtungen nach §17d KHG*. Institut für das Entgeltsystem im Krankenhaus: Siegburg.

Jencks SF, Horgan C, Taube CA (1987). Evidence on provider response to prospective payment. *Medical Care* **25**, 37–41.

Jensen DD, Cohen PR (2000). Multiple comparisons in induction algorithms. *Machine Learning* **38**, 309–338.

Lave JR (2003). Developing a Medicare prospective payment system for inpatient psychiatric care. *Health Affairs* **22**, 97–109.

Lave JR & Frank RG (1990). Effect of the structure of hospital payment on length of stay. *Health Services Research* **25**, 327–347.

McCrone P (1994). Allocating resources in psychiatric hospitals according to casemix. *Psychiatric Bulletin* **18**, 212–213.

McCrone P (1995). Predicting mental health service use: diagnosis based systems and alternatives. *Journal of Mental Health* **4**, 31–40.

McCrone P, Phelan M (1994). Diagnosis and length of psychiatric in-patient stay. *Psychological Medicine* **24**, 1025–1030.

McCrone P, Leese M, Thornicroft G, Schene A, Knudsen HC, Vázquez-Barquero JL, Tansella M, Becker T (2001). A comparison of needs of patients with schizophrenia in five European countries: the EPSILON Study. *Acta Psychiatrica Scandinavica* **103**, 370–379.

Moreno K (2007). Adaptation of activity-based-costing (ABC) to calculate unit costs in mental health care in Spain. *European Journal of Psychiatry* **21**, 117–123.

Payment by Results (2012). *NHS Costing Manual*. Department of Health: London.

Rupp A, Steinwachs DM, Salkever DS (1984). The effect of hospital payment methods on the pattern and cost of mental health care. *Hospital and Community Psychiatry* **35**, 456–459.

Suarez M, Bravo-Ortiz M, Fernandez-Liria A, Gonzalez-Juarez C (2011). Effectiveness of Continuity-of-Care Programs to reduce time in hospital in persons with schizophrenia. *Epidemiology and Psychiatric Sciences* **20**, 65–72.

Tulloch AD, Fearon P, David AS (2011). Length of stay of general psychiatric inpatients in the United States:

systematic review. *Administration and Policy in Mental Health* **38**, 155–168.

Wahlbeck K (2011). European comparisons between mental health services. *Epidemiology and Psychiatric Sciences* **20**, 15–18.

Weinmann S, Koesters M, Becker T (2007). Effects of implementation of psychiatric guidelines on provider performance and patient outcome: systematic review. *Acta Psychiatrica Scandinavica* **115**, 420–33.

Yamauchi K (1997*a*). Designing a new payment system for psychiatric care – developing of a case-mix classification system. *Journal of the Japanese Society on Hospital Administration* **34**, 155–200.

Yamauchi K (1997*b*). Designing a new payment system for psychiatric care – relationship between the costs on inpatient care and the revenue from insurances. *Journal of the Japanese Society on Hospital Administration* **34**, 60–71.