

## Developing and comparing methods for measuring the content of care in mental health services

B. Lloyd-Evans · M. Slade · D. P. Osborn ·  
R. Skinner · S. Johnson

Received: 24 June 2009 / Accepted: 25 January 2010 / Published online: 14 February 2010  
© Springer-Verlag 2010

### Abstract

**Background** Measurement of what happens in mental health services is needed to describe services, identify variation in care provision and understand service outcomes. However, there is no consensus about appropriate methods or measures for content of care. Previous research has primarily used a single information source and prioritised staff over patient perspectives on content of care. This study aims to enhance understanding of how to measure content of care by developing and evaluating four instruments, each using a different measurement method.

**Methods** Development is described by four instruments—CaSPAR, CaRICE, CCCQ-S and CCCQ-P—which use researcher observation, staff-report and patient-report to

measure the intensity and nature of care at services. Inter-rater reliability of CaRICE and CCCQ-S was investigated. Concordance between staff and patient perspectives was explored through assessing inter-rater agreement of CCCQ-S and CCCQ-P questionnaires. The convergence of data from the measures collected in an inpatient multi-site study was investigated.

**Results** CaRICE demonstrated good inter-rater reliability ( $\kappa = 0.71$ ). CCCQ-S inter-rater reliability was poor. Concordance between staff and patient reports was low: there was a trend for patients to report less care received than staff reported had been provided. Results from CaSPAR, CaRICE and CCCQ-P exhibited divergence, indicating possible differences in patient, staff and researcher perspectives.

**Discussion** Information about content of care should be sought as close as possible to the point of delivery. There may be differing, valid perspectives about care provided by services. Further development of methods to measure content of care is required. Meanwhile, a multi-methods approach should be adopted, which allows inclusion of different perspectives (specifically including the patient's perspective) and triangulation of results from different measurement methods. CaSPAR, CaRICE and CCCQ-P can provide multi-perspective content of care measurement in inpatient services.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00127-010-0192-4) contains supplementary material, which is available to authorized users.

B. Lloyd-Evans (✉) · S. Johnson  
Department of Mental Health Sciences, University College  
London (Bloomsbury Campus), Charles Bell House,  
67-73 Riding House Street, London W1W 7EJ, UK  
e-mail: b.lloyd-evans@ucl.ac.uk

B. Lloyd-Evans · S. Johnson  
Camden and Islington NHS Foundation Trust,  
Early Intervention Service, 3rd Floor Bedford House,  
125-133 Camden High Street, London NW1 7JR, UK

M. Slade · R. Skinner  
Health Service and Population Research Department,  
Institute of Psychiatry, Kings College London,  
Denmark Hill, London SE5 8AF, UK

D. P. Osborn  
Department of Mental Health Sciences,  
Royal Free Hospital, University College London  
(Hampstead Campus), Pond Street, London NW3 2QG, UK

**Keywords** Process measurement · Content of care ·  
Inpatient · Mental health services

### Background

Measurement of content of care is necessary to describe what happens in services [4], to assess variation in service

delivery between similarly labelled services [6] and to understand service outcomes [15, 21]. Quantitative measurement of care delivered is a prerequisite for establishing associations between the process of care and outcomes and an empirical basis for service models or quality indicators [11, 29]. Understanding the mechanism by which service outcomes are achieved is particularly important for complex interventions [8], where the active ingredients may not be self-evident.

There is therefore a need to develop methods to measure the content of care in mental health services, particularly for acute inpatient services whose purpose and content have been described as insufficiently defined [10] and where little is known about what happens to patients during an admission [24]. However, a recent review of measures of content of care [19] concluded that there is no consensus about ideal methods or measures of service content in mental health services. Existing measures do not fully meet challenges to effective measurement of content of care by demonstrating robust psychometric properties, providing adequate depth of information about the nature of service interventions, demonstrating feasibility in a range of service settings and including both staff and patient perspectives about care provided.

A lack of adequate measures has been identified in particular to describe the types of care provided in acute inpatient services [19], where the large number of caregivers and acute illness of care-recipients add to the challenge of establishing what care is provided to patients. The most common approach to assessing the care provided in inpatient services has been using researcher observation to quantify levels of staff–patient interaction on wards, developing methods pioneered by Altschul [2]. A staff–patient contact in inpatient services can be reliably established through observation [23, 25, 26, 30] but previous studies have encountered two further challenges to measuring content of care using observation. The first is obtaining representative data. Sanson-Fisher et al. [26] reported observing all staff–patient interactions in a study of inpatient services. However, this may not be achievable in modern acute wards, where not all areas are immediately visible to an observer, substantial amounts of care are provided off the ward and ethical considerations may limit the extent of observations. More recent studies have used three approaches to address this problem: limiting observation to communal areas of a ward [9]; recording staff activity without recording which or how many patients receive attention [14, 30]; or recording only the activity of a small number of consenting service users [14]. None of these approaches guarantees wholly representative information about how much contact with staff patients receive.

The second challenge to assessing the content of care using researcher observation is understanding what care is

being provided during observed staff–patient interactions. Previous studies required observers to describe inpatient staff activity using 23 [7] or 35 [14] categories, without establishing whether this could be done reliably. Paul [23] demonstrated very good reliability for coding observed staff behaviour during interactions with patients into 21 categories. These mainly relate to the style of care (positive or negative, verbal or non-verbal) but provide some information about types of care, distinguishing group or individual contact and whether physical force was used. Observation measures have however not demonstrated the ability to reliably provide depth of information about the purpose of contacts or types of interventions provided.

This study aims to increase understanding about how to achieve representative and informative measurement of the content of care in mental health services. The development of four instruments for use in inpatient settings is described. Their psychometric properties are explored and results from each instrument are compared, using data collected for The Alternatives Study [17, 18], a national multi-site study of acute inpatient services. Conclusions about how to measure content of care in mental health services are presented and needs for further research identified.

## Methods

### Instrument development

Four instruments were developed: Camden Staff–Patient Activity Record (CaSPAR); Camden Record of Inpatient Care Events (CaRICE); Camden Content of Care Questionnaire—staff version (CCCQ-S); Camden Content of Care Questionnaire—patient version (CCCQ-P).

Our intention in developing four instruments was to triangulate results from different information sources and data collection methods. Collectively, the instruments provide information about the intensity of staff–patient contact and the types of care delivered at services. All measure only direct patient care, allowing comparison of similar variables from each instrument. Choice of information source and measurement methods for the instruments was informed by a review of content of care measures [19]. Three sources informed the choice of categories of care used in three of the instruments to describe the nature of care provided at services: qualitative research asking staff at an inpatient ward and a crisis house to describe their activities [16]; a Delphi process with community mental health professionals categorising their activities [12]; and consultation with the steering group of The Alternatives Study [17], comprising service users, carer representatives, clinicians and researchers with relevant experience.

## Measures

*CaSPAR* is a researcher-rated momentary time recording measure based on researcher observation supplemented with staff information. At 28 prespecified time points, including evening and weekends, each patient is recorded as on or off the unit and as being in active engagement with staff (non-verbal—such as silently playing a board game or under restraint or verbal) or not engaged. Data from each recording point can be aggregated to provide a *CaSPAR* total care score for each service, representing the mean proportion of patients observed in contact with staff.

*CaSPAR* uses direct observation by a researcher as a primary information source, supplemented by additional information from staff (e.g. about patients being escorted by staff off the unit) as required to provide complete information about staff–patient contact at recording points. In comparison to wholly observation-based measures of content of care, *CaSPAR*, by using two information sources (observation and staff-report), seeks to maximise the representativeness of data collected and thus the validity of the measure, at the risk of compromising reliability.

*CaRICE* is an event-recording measure using staff report. All staff at a service are asked to complete *CaRICE* forms each day during recording periods. Staff record each contact with a patient lasting 5 min or more, noting its duration and describing the care provided using one or more of 21 categories provided. Abbreviated description of the categories and four subscales into which they can be placed are provided in Table 1.

Data from each completed recording form can be summed to provide a total service score for each day. These data can be aggregated to produce one *CaRICE* total care

score for each service, representing the mean minutes of staff contact provided per patient per day. Subscale and individual category scores can be similarly generated. Service data for the mean proportion of staff time spent in contact with patients can also be calculated.

*CaRICE* gives a staff perspective about care provided. The use of staff report allows data about the nature of care to be collected which researchers might not be able to infer from brief, non-intrusive observation. The 21 categories used in *CaRICE* (Table 1) provide greater depth of information about the types of care at services than previous event recording measures, which have used between 5 and 11 categories [19]. *CaRICE* reduces risk of recall bias because staff activity is recorded at or very soon after the time it happens. The representativeness of data is maximised by inclusion of all staff working during recording periods. The minimum threshold of 5 min for recorded contacts between staff and patients enhances the feasibility of *CaRICE* for inpatient settings because it does not require staff to record the potentially numerous, very brief contacts they may have with patients (such as greeting patients in passing). The *CaRICE* results presented in this paper were collected from a one-week recording period. This was chosen as the minimum time period required to capture all routine activities occurring at services and thus provide representative data.

*CCCQ-S* is a retrospective questionnaire about all the care received by individual patients during an inpatient admission completed by a single staff respondent. *CCCQ-S* uses the same 21 categories of care as *CaRICE*: for each category, information is provided about whether and how frequently (using a seven-point ordinal scale) the patient received help from staff during his admission. Summing

**Table 1** Categories of care used in *CaRICE*, *CCCQ-P* and *CCCQ-S*

Social interventions subscale		Psychological interventions subscale		Physical and pharmacological interventions subscale		General care organisation subscale	
Item #	Description	Item #	Description	Item #	Description	Item #	Description
1	Help with housing	9	Symptom coping strategies	14	Medication information or discussion	8	Assessment
2	Help with finances	10	Help with relationships	15	Medication review	20	Care planning meeting
3	Help with legal matters	11	Help with distressing past events	16	Medication dispensing or practical help	21	Care coordination
4	Help with current activity (while an inpatient)	12	Help with drug or alcohol problems	17	Physical health care		
5	Help with future activity (following discharge)	13	Illness education	18	Safety: observation		
6	Help with activities of daily living			19	Safety: restraint		
7	Carer/family support						

responses from each category about whether care was provided generates a CCCQ-S range of care score between 0 and 21. Summing responses from each category about the frequency with which care was received generates a CCCQ-S total care score from 0 to 147, reflecting the intensity of care provided during a patient's admission.

CCCQ-S is potentially useful because, unlike CaSPAR and CaRICE, it provides information about care provided to specific patients. By seeking retrospective information from a single respondent, CCCQ-S provides a low-burden way to obtain information about care provided to individual patients over a whole admission. Because mental health services do routinely identify a care coordinator, primary nurse or key worker to oversee and coordinate a patient's care during their admission, one respondent was thought likely to possess the required information.

CCCQ-P is a retrospective questionnaire similar to CCCQ-S but completed by the patient. Like CCCQ-S, it generates a range of care score between 0 and 21 and a CCCQ-P total care score between 0 and 147. It provides information about content of care from a third information source (in addition to staff-report and researcher observation) and allows direct comparison with similar CCCQ-S data of staff and patient perspectives on care provided at services. By seeking information at the time of discharge, it maximises the likelihood of patients being well enough to consent and complete the measure and allows assessment of the care received during a whole inpatient stay.

#### Psychometric properties of instruments

Four psychometric investigations of the instruments described above were conducted. Three of these investigations involve data about the content of care in inpatient services collected from Phase 2 of The Alternatives Study, described elsewhere [17].

#### *Inter-rater reliability of CaRICE*

**Procedures:** 21 vignettes describing contact between staff and patients on acute wards were created from anonymised case note entries. Respondents were asked to code each vignette using the categories of care provided in CaRICE.

**Sample:** vignettes were coded by a pragmatically chosen sample of clinicians from Camden and Islington ( $n = 21$ ), including psychologists, psychiatrists, nurses and social workers.

**Analysis:** for each vignette, the number of clinicians using each CaRICE category to rate the vignette was entered into a database using Stata software and a  $\kappa$

value was calculated to assess overall inter-rater reliability.

#### *Concordance between CCCQ-S and CCCQ-P*

**Setting:** data were collected from 8 inpatient services. These comprised three non-hospital crisis houses (Services 1–3) and five inpatient acute wards (Services 5–8). Service 1 was a 9-bedded crisis house for Black Minority Ethnic patients in East London, run by a voluntary sector Housing Association and staffed by non-clinical social care workers. Service 2 was an 8-bedded residential unit within a Community Mental Health Resource Centre in Staffordshire, staffed similarly to an acute ward predominantly by nursing and medical staff. Service 3 was a 4-bedded crisis house in North East England, run by the local Crisis Resolution Team and situated within a larger social services rehabilitation hostel. Services 4–8 were all general acute admission inpatient services ranging between 15 and 25 beds, situated in East Birmingham, East London, Staffordshire, North East England and South Birmingham respectively.

**Sample:** both CCCQ-S and CCCQ-P forms were obtained for 108 patients.

**Procedures:** CCCQ-S and CCCQ-P questionnaires were completed by staff and patients respectively, at the time of the patient's discharge from the service.

**Analysis:** data were entered in an electronic database using Stata software. Concordance between staff and patient respondents was assessed: from staff and patient responses for each patient, for each category of care a  $\kappa$  value was calculated to assess the inter-rater reliability of binary data about whether any care in that category had been provided; Spearman's rho was calculated for ordinal data about the frequency of provision of care.

#### *Inter-rater reliability of CCCQ-S*

**Services:** data were collected from two of the services described above: a crisis house in North East England and an acute inpatient ward in South Birmingham.

**Sample:** two CCCQ-S forms were obtained for 46 patients.

**Procedures:** at two inpatient services (an acute ward and a community-based crisis house), two members of staff who had worked closely with the patient were identified and asked to complete CCCQ-S questionnaires at the time of the patient's discharge.

**Analysis:** data were entered in an electronic database using Stata software and analysed similarly to the

concordance exploration of CCCQ-S and CCCQ-P described above.

### Convergent validity of CaRICE and CCCQ-P

Setting and procedures: CaSPAR, CaRICE and CCCQ-P data were collected from the 8 inpatient services described above.

Sample: 28 CaSPAR recordings were made at each service ( $n = 224$ ). One week of CaRICE recording was carried out at each service. CCCQ-P questionnaires were completed with 314 patients.

Analysis: the convergence of CaRICE and CCCQ-P results with results from CaSPAR assessed through comparison of descriptive data and clustering of services indicated by Duncan's multiple range tests (a test of homogeneity in subsets, exploring whether and what subgroups can be identified within datasets). CaSPAR, CaRICE and CCCQ-P total scores all relate to the intensity of direct care provided at services. CaSPAR data were used as a benchmark against which to compare CaRICE and CCCQ-P because CaSPAR primarily used momentary time recording based on researcher observation, which has previously demonstrated good reliability [23, 30].

## Results

CaSPAR, CaRICE, CCCQ-S and CCCQ-P data used in the psychometric explorations presented here were collected from 8 acute inpatient services. In 224 CaSPAR recordings, the status of patients (with staff or not) could be ascertained for 99% of patients. In 5 days of CaRICE recording at each service, CaRICE forms were completed by staff from 871 out of 919 shifts, a response rate of 94.7%. Completed CCCQ-P forms were obtained from 314 respondents from 447 patients approached, a response rate of 70.2%. Staffs were asked to complete CCCQ-S forms for 465 patients: 433 forms were completed, a response rate of 93.1%.

### Inter-rater reliability of CaRICE

Responses of 21 clinicians categorising care described in 21 vignettes are presented in Table 2.

Analysis produced a  $\kappa$  value for inter-rater agreement of 0.71. This represents good inter-rater reliability, being between 0.61 and 0.8 [1]. In 20 out of 21 vignettes, over 80% of respondents used the same single category of care to describe staff activity. Fewer than 20% of respondents used any other single category of care in 17 out of the 21 vignettes. Assessment was the most widely used category,

used by at least one respondent in 9 out of the 21 vignettes.

### Concordance between CCCQ-S and CCCQ-P

Table 3 presents analyses of concordance between staff and patient report of care received by 108 patients, based on CCCQ-S and CCCQ-P data.

A  $\kappa$  score of more than 0.40, indicating moderate agreement between respondents [1] was achieved between staff and patient respondents about whether care had been received in only 3 out of 21 categories and about how frequently care had been received (Spearman's rho > 0.40) in only 2 out of 21 categories. Concordance between CCCQ-S and CCCQ-P data was thus very low. There was a trend for patients to report less care as received than staff reported as provided: fewer patients than staff said care had been received in 17 out of 21 categories.

### Inter-rater reliability of CCCQ-S

Table 4 presents analyses of inter-rater reliability of CCCQ-S data from two staffs completed for 46 patients.

Moderate inter-rater agreement ( $\kappa > 0.40$ ) about whether patients had received any care of each type was achieved for only 5 out of 21 categories. Moderate agreement (Spearman's rho > 0.40 [1]) about the frequency of care provided was achieved for 6 out of 21 categories. Overall, CCCQ-S therefore demonstrated poor inter-rater reliability.

Secondary analysis (provided in the online data supplement to this article—Table S1) indicated that inter-rater agreement was no better among respondents at the community-based crisis house than the inpatient ward, despite patient admissions being substantially shorter. While recall bias is a potential source of unreliability for CCCQ-S and CCCQ-P because the measures gather information about care received retrospectively at the point of a patient's discharge, this investigation does not suggest that the time period over which CCCQ-S data are collected is the major determinant of its poor reliability.

### Convergent validity of CaRICE and CCCQ-P

CaSPAR, CaRICE and CCCQ-P total scores from 8 services are presented in Table 5, with each column ordered starting with the service providing most care.

Divergence in the rankings of services on different measures is evident. For example, Service 3 scores highest on CaRICE but lowest on CaSPAR; Service 8 scores lowest on CaRICE but highest on CCCQ-P. A Duncan's multiple range test of CaSPAR data indicated that Service

**Table 2** CaRICE inter-rater reliability test: coding of vignettes by clinicians ( $n = 21$ )

Vignette	Category of care																					
	Housing	Finances	Legal	Current activity	Future activity	Activities of daily living	Family support	Assessment	Symptom coping strategies	Relation ships	Past events	Drugs and alcohol	Illness education	Medication information	Medication review	Medication practical help	Physical health care	Observations	Restraint	Care planning meeting	Aftercare coordination	
1	0	0	0	0	2	11	0	0	2	0	0	0	0	0	0	0	14	0	0	0	0	
2	0	0	0	18	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	19	4	0	0	0	0	0	0	0	0	0	1	0	1	0
4	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	3	19
5	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	3	18	0	0
6	0	0	0	0	0	0	0	0	0	0	0	1	0	18	4	4	0	0	0	0	2	0
7	0	0	0	3	0	0	0	1	20	0	0	0	0	0	0	0	0	0	1	0	0	0
8	0	0	0	6	1	0	21	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0
10	0	0	0	0	1	0	0	5	4	0	1	0	19	0	1	0	0	0	0	0	0	2
11	0	0	0	0	0	21	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	2	0	0	0	21	0	0	1	0	1	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	3	0	1	16	1	0	0	0	0	1	0
14	0	0	21	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	7	2	0	0	0	0	0	1	0	0	0	19	0	0	1
16	0	0	0	4	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
17	0	0	0	0	0	0	3	0	0	21	1	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	2	9	1	17	0	0	0	0	0	0	0	0	0	0	0
19	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	3
21	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2



**Table 3** CCCQ-S and CCCQ-P concordance ( $n = 108$ )

	Was care received?			How frequently? (ordinal scale 0–7)		
	Patient Yes	Staff Yes	$\kappa$	Spearman's rho	Patient median	Staff median
1. Housing	24	35	0.332	0.360	0	0
2. Finances	20	30	0.280	0.299	0	0
3. Legal	14	12	0.301	0.315	0	0
4. Current activity	48	56	−0.070	0.003	0	1
5. Future activity	33	51	0.054	0.023	0	0
6. Activities of daily living	19	33	0.104	0.102	0	0
7. Family	49	73	0.238	0.167	0	3
8. Assessment	77	80	0.021	0.156	3	4
9. Coping strategies	46	88	−0.016	0.041	0	4
10. Relationships	15	35	0.250	0.346	0	0
11. Past events	34	48	−0.010	−0.067	0	0
12. Drugs alcohol	21	43	0.404	0.473	0	0
13. Illness education	45	59	0.052	−0.018	0	1
14. Medication concordance	49	63	0.153	0.237	0	1
15. Medication review	57	58	0.424	0.396	1	1
16. Medication practical	87	63	0.134	0.128	7	5
17. Physical health care	50	32	0.155	0.178	0	0
18. Observations	34	29	0.084	0.060	0	0
19. Restraint	7	12	0.484	0.510	0	0
20. Care planning	58	76	0.264	0.242	1	3
21. Aftercare coordination	56	81	0.086	−0.026	1	1

2 provided significantly more staff–patient contact than a cluster containing all seven other services. This was not replicated in Duncan's tests of CaRICE and CCCQ-P data, where no comparably clear clustering of services emerged. Comparison with CaSPAR data in this multi-site study does not establish convergent validity for CaRICE and CCCQ-P.

## Discussion

### Main findings

Four main findings can be drawn from the psychometric exploration of instruments described above. First, patient and staff reports of care provided at inpatient services, measured by CCCQ-S and CCCQ-P, exhibit poor concordance. The extent to which this is a result of scale error in imprecision or how far it represents true disagreement between staff and patient perspectives is not resolved by this investigation. However, the trend suggested in the data (in 17 out of 21 categories) for patients to report receiving less care than staff think has been provided suggests there may be systematic differences in

staff and patient perspectives on care provided at services.

Second, the inter-rater reliability of CCCQ-S is unacceptably low. Psychometric explorations indicate that neither recall bias nor problems of using the 21 categories to describe care are the primary source of this observed unreliability. The poor inter-rater reliability results for CCCQ-S occurred despite care being taken to recruit two staff respondents who both knew the patient well and felt able to report what care had been provided. It therefore seems probable that individual staff members, even those like primary nurses who are specifically assigned to a patient, do not have an overall awareness of all the care provided to patients. This may reflect the complex nature of inpatient care, provided by multiple care-givers. The use of CCCQ-S as a measure of content of care is not supported by results from this study.

Third, there is equivocal evidence for the psychometric robustness of CaRICE. The good inter-rater reliability demonstrated in testing using case note vignettes provides some support for the use of CaRICE as a measure of content of care. However, divergence of CaRICE results from those generated by CaSPAR, a primarily observation-based measure, suggests possible unreliability of CaRICE.

**Table 4** CCCQ-S Inter-rater reliability ( $n = 46$ )

	Was care received?			How frequently? (ordinal scale 0–7)		
	Rater 1 Yes	Rater 2 Yes	$\kappa$	Spearman's Rho	Rater 1 median	Rater 2 median
1. Housing	14	11	0.618	0.681	0	0
2. Finances	12	13	0.506	0.531	0	0
3. Legal	11	4	0.312	0.348	0	0
4. Current activity	13	26	0.401	0.352	2	3
5. Future activity	24	22	0.129	0.179	0	0
6. Activities of daily living	32	16	0.605	0.603	0	0
7. Family support	10	28	0.207	0.307	2	2
8. Assessment	43	40	0.148	−0.047	4	4
9. Coping strategies	37	42	−0.137	0.189	5	5
10. Relationships	19	17	0.362	0.336	0	0
11. Past events	19	19	0.191	0.191	0	0
12. Drugs or alcohol	23	19	0.652	0.661	0	1
13. Illness education	32	28	0.335	0.578	1	3
14. Medication concordance	35	30	0.328	0.386	2	3
15. Medication review	30	26	0.276	0.127	1	1
16. Medication practical	33	29	0.313	0.399	2	6
17. Physical health care	16	9	0.306	0.415	0	0
18. Observations	19	13	0.247	0.238	0	0
19. Restraint	2	4	0.292	0.329	0	0
20. Care planning	35	24	0.244	0.379	1	1
21. Aftercare coordination	37	40	−0.027	−0.085	1	1

**Table 5** CaSPAR, CaRICE and CCCQ-P: comparison of results from 8 services

CaSPAR (224 recordings) Service mean proportions of patients in contact with staff	CaRICE (1 week recording per service) Service mean minutes of contact per patient per day	CCCQ-P ( $n = 314$ ) Service mean total scores
Service 2 21.6%	Service 3 160.4	Service 8 34.4
Service 7 14.1%	Service 6 154.1	Service 4 31.9
Service 6 13.5%	Service 2 139.5	Service 6 31.8
Service 4 10.8%	Service 1 133.3	Service 5 30.6
Service 8 8.8%	Service 7 131.7	Service 2 29.8
Service 1 8.6%	Service 5 121.7	Service 7 29.4
Service 5 8.0%	Service 4 109.6	Service 3 29.1
Service 3 4.6%	Service 8 82.6	Service 1 20.6

Results may also be influenced by differences in the variables measured by the two instruments (for example, contacts of less than 5-min duration are included in

CaSPAR but not in CaRICE) and/or differences in staff and researcher perspectives about what constitutes staff–patient contact.



Fourth, gathering information about the content of care in acute inpatient services is feasible using all four of the instruments described in this paper. Staff on busy acute wards will provide contemporaneous information about their activity and can identify the whereabouts of staff and patients. Patients can be recruited and give informed consent at the end of an acute admission and provide information about care received.

#### Strengths and limitations of the study

This study's incomplete testing of the psychometric properties of instruments is a limitation. Reliability testing of CaRICE was not possible in actual clinical settings: because much inpatient care is provided by single staff members, only one description of care provided can be gathered; because the measure is designed to be completed contemporaneously, test–retest explorations are not possible. Vignettes have been used in previous studies to test reliability [4]; however, the divergence of results gathered from CaRICE and CaSPAR casts doubt on the reliability of CaRICE and the adequacy of reliability testing using vignettes. It is possible that descriptions of care in case notes focus on concrete interventions and do not include situations where contact with patients or its nature is more ambiguous. The inter-rater reliability testing of CaRICE is therefore not conclusive. Test–retest reliability, of CaRICE using vignettes or of CCCQ in actual clinical settings, was not undertaken because of resource constraints. The additional gains from exploring test–retest reliability were limited because CCCQ-S had already been demonstrated to be unreliable and because of the limitations of using vignettes. However, test–retest explorations would have provided some further information about the reliability of CaRICE and sources of unreliability of CCCQ-S and could usefully be explored in future.

The theoretical basis for, and refinement through consultation and piloting of, the descriptive categories used in CaRICE and CCCQ-P suggests that they may allow respondents to describe the content of staff–patient interactions adequately. The good response and completion rates obtained for CaSPAR, CaRICE and CCCQ-P in this investigation provide some confidence that representative data about service provision can be obtained from the measures. However, other than the exploration of convergence described in this paper, formal tests of the measures' validity have not been undertaken. Establishing convergent or predictive validity for instruments in this study is problematic, given the lack of directly comparable instruments with demonstrated psychometric properties and the lack of knowledge about how service content in inpatient services affects outcomes.

The uncertainty about the psychometric properties of CaSPAR, CaRICE and CCCQ-P makes interpretation of divergence in results from the measures problematic. Possible explanations include psychometric inadequacies, but also the sampling, differences in the specific variables measured or perspectives of respondent groups. The one-week sampling period used for CaRICE, for example, may have coincided with unusual circumstances at some services and been insufficient to reflect typical care provision. Services which provide the most frequent interventions may not have also provided the greatest duration of staff contact with patients. Patients, staff and researchers may all have interpreted staff behaviour differently: for instance, a member of staff detailed to an hour of close observation with a patient may record an hour of patient contact using CaRICE; a researcher may or may not witness interaction with the patient at the time of a CaSPAR recording; the patient might not experience receiving any care at all. This study, however, cannot identify how or how much these factors may have influenced results from each instrument.

The instruments described in this study were developed specifically to assess the content of care in acute inpatient settings. The categorisation of care in CaRICE and CCCQ may not adequately describe the care provided in other settings, such as nursing homes or non-acute residential care. A lack of instruments to assess the nature of care in inpatient services has been identified [19]. This study has developed instruments to address this need and demonstrated their ability to provide adequately complete data in inpatient settings. It has also increased broader understanding of appropriate methods to measure service content in mental health settings in the following ways:

- (i) Patient perspectives have seldom been included in measuring service content [19]. The possibility of patients, staff and researchers having different views about what constitutes care or its nature have been largely ignored in previous studies assessing service content. Where addressed, reconciliation of differences has been sought [13], without explanation of how this might be achieved. This study is innovative in directly comparing data on content of care gathered from different information sources and finding that patients, staff and researchers may indeed have different perspectives.
- (ii) The completeness of data gathered from CaRICE and CCCQ-P indicates that staff and patients were able to use 21 categories to describe care provided. The inter-rater reliability testing of CaRICE produced a  $\kappa$  value comparable to those generated by Brekke [4] testing the Daily Contact Log, a staff-report measure categorising care into seven types. (Brekke reports  $\kappa$  values of 0.59 and 0.68 in two tests using vignettes

[4]). This suggests that increasing the number of categories used to describe care may not jeopardise the reliability of information obtained. The instruments described in this study and the categorisation of care used in CaRICE and CCCQ-P can provide greater depth of information can be obtained about the nature of inpatient service content than was previously available.

- (iii) This study explored whether inpatient staff can provide an overview of care given to a patient during an admission and provide information about interventions delivered by colleagues. It has established that this cannot be provided with adequate accuracy. This negative finding can inform future choice of method in measuring the content of inpatient care.

### Research implications

Given current knowledge, two recommendations can be made for how to measure content of care in mental health services. First, a multi-method approach should be used. This study suggests that care may be a construct which defies objective definition. It may be better understood, like ward atmosphere [22], patient needs [28] or therapeutic alliance [20], as a subjectively experienced phenomenon about which there are differing, valid perspectives. Further research is required to understand the extent and nature of any such differing perspectives. The development of a single, multi-perspective instrument with adequate reliability is desirable, which can capture both staff and patient perspectives about care provided. Meanwhile, multi-method assessment can include different perspectives and account to an extent for the uncertain reliability of instruments by triangulating data from more than one source. Features of, or differences between, services which are found repeatedly and by more than one measurement method can be accepted with greater confidence than if based on data from one measure alone.

Second, measures should seek information as close, temporally and physically, to the object of measurement as possible. Staff and patients should only be asked about their own experience of giving or receiving care. Information about interventions should be obtained from staff at or very near the time they are delivered. Given the uncertain reliability of staff and patient-report measures, direct observation by researchers, which has demonstrated reliability [3, 9, 23, 25–27, 30], should be included in assessment of service content, where possible and can be used to compare with data from other information sources.

Future research could increase understanding of how to measure content of care and address some of the limitations of this study in the following three ways:

- (i) Collecting data from services using all the instruments described in this study over the same brief time period, e.g. a day or a week, could illuminate causes of divergence of results from different measures. If results were as divergent as those reported in Table 5, inherent features of the measures—respondents' differing perspectives, differences in the variables measured, or unreliability—would appear to cause divergence. More convergent results would suggest that divergence was substantially influenced by fluctuations in care provided at services over short periods of time. Longer time periods for data collection, consistent across instruments, would then be indicated for assessing service content.
- (ii) Differences in perspective between staff, patient and researcher perspectives could be further explored. A study could be designed requiring participants to watch video footage of real or enacted contact between staff and patients and identify and describe interactions using CaRICE categories. This would allow investigation of inter and intra-group differences in perspective about what constitutes contact between staff and patients and the nature of care being provided.
- (iii) The extent of information about service content which can reliably be obtained by researcher-observation could be explored. Broad distinctions about the quality of staff–patient contacts, as positive, negative or neutral can be reliably made by observers [27, 30]. Paul [23] demonstrated more nuanced distinctions about the style of care provided in staff–patient interactions can also be made. Whether further categorisation of the content of service interventions can be provided reliably by observers could be a focus for future research.

### Conclusion

Despite repeated calls for investigation of what happens in services and how it impacts on service outcomes [5, 6, 8, 11, 15, 29], process measurement in mental health services has been under-researched. No gold standard measures of content of care have been established and findings are rarely confirmed by more than one study [19]. The considerable methodological difficulties in measuring service content may help to explain this lack of attention. Focus by service audit on easily measured specifics of care, such as medication prescription or CPA documentation, is understandable. Nevertheless, the rationale remains strong for seeking to measure all service content, to understand what is provided and differentiate services.

Meanwhile, multi-method assessment of content of care and triangulation of results from different information sources is desirable. This study describes instruments with which this can be undertaken in inpatient services, to assist the goals of describing and distinguishing services and understanding service outcomes.

**Acknowledgments** We are very grateful to the people using, working in and managing the participating services who gave their time generously, and for advice and input from the steering group for The Alternatives Study. This report presents independent research commissioned by the National Institute for Health Research Service Delivery and Organisation R&D Programme. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. The authors have no competing interests to declare.

## References

- Altman D (1991) Practical statistics for medical research. Chapman and Hall, London
- Altschul A (1972) Patient–nurse interaction: a study of interaction patterns in acute psychiatric wards. Churchill Livingstone, Edinburgh
- Bowie P, Mountain G (1993) Using direct observation to record the behaviour of long-stay patients with dementia. *Int J Geriatr Psychiatry* 8:857–864
- Brekke J (1987) The model-guided method for monitoring program implementation. *Eval Rev* 11(3):281–299
- Brugha TS, Lindsay F (1996) Quality of mental health service care: the forgotten pathway from process to outcome. *Soc Psychiatry Psychiatr Epidemiol* 31(2):89–98
- Burns T, Priebe S (1996) Mental health care systems and their characteristics: a proposal. *Acta Psychiatr Scand* 94(6):381–385
- Cormack D (1976) Psychiatric nursing observed. Royal College of Nursing, London
- Craig P, Dieppe P, Macintyre S, Mitchie S, Nazareth I, Petticrew M (2008) Developing and evaluating complex interventions: the new Medical Research Council guidance. *Br Med J* 337:979–983
- Dean R, Proudfoot R (1993) The quality of interactions schedule (QUIS): development, reliability and use in the evaluation of two domus units. *Int J Geriatr Psychiatry* 8:819–826
- Department of Health (2005) Mental Health Policy Implementation Guide: Adult Acute Inpatient Care Provision. Department of Health
- Donabedian A (1980) Explorations in quality assessment and monitoring. The definition of quality and approaches to its assessment, vol 1. Health Administration Press
- Fiander M, Burns T (2000) A Delphi approach to describing service models of community mental health practice. *Psychiatr Serv* 51(5):656–658
- Glick I, Burti L, Suzuki K, Sacks M (1991) Effectiveness in psychiatric care: 1. A cross-national study of the process of treatment and outcomes of major depressive disorder. *J Nerv Mental Dis* 179(2):55–63
- Higgins R, Hurst K, Wistow G (1999) Nursing acute psychiatric patients: a quantitative and qualitative study. *J Adv Nurs* 29(1):52–63
- Johnson S, Salvador-Carulla L (1998) Description and classification of mental health services: a European perspective. *Eur Psychiatry* 13:333–341
- Johnson S, Billings J, Bingham C, Pilling S, Bebbington P, Morant N, Dalton J, McNicholas S (2001) An evaluation of Drayton Park Women’s Crisis Service, a community-based alternative to the psychiatric ward. Unpublished report
- Johnson S, Gilbert H, Lloyd-Evans B, Slade M (2007) Acute inpatient psychiatry: residential alternatives to hospital admission. *Psychiatr Bull* 31:262–264
- Johnson S, Gilbert H, Lloyd-Evans B, Boardman H, Leese M, Osborn D, Shepherd G, Thornicroft G, Slade M (2009) Alternatives to hospital admission: emerging models and populations served. *Br J Psychiatry* 194(5):456–463
- Lloyd-Evans B, Slade M, Johnson S (2007) Assessing the content of mental health services: a review of measures. *Soc Psychiatry Psychiatr Epidemiol* 42(8):673–682
- McCabe R, Roeder-Wanner U, Hoffman K, Priebe S (1999) Therapeutic relationships and quality of life: associations of two subjective constructs in schizophrenia patients. *Int J Soc Psychiatry* 45:276–283
- Mechanic D (1996) Emerging issues in international mental health services research. *Psychiatr Serv* 47(4):371–375
- Moos R (1996) Ward atmosphere scale manual, 3rd edn. Mind Garden
- Paul G (1987) Rational operations in residential treatment settings through ongoing assessment of staff and client functioning. In: Peterson D, Fishman D (eds) *Assessment for Decision*. Rutgers University Press, pp 145–203
- Quirk A, Lelliot P (2001) What do we know about life on acute psychiatric wards in the UK? A review of the research evidence. *Soc Sci Med* 53(12):1565–1574
- Sandford D, Elzinger R, Iverson R (1990) A quantitative study of nursing staff interactions in psychiatric wards. *Acta Psychiatr Scand* 81:46–51
- Sanson-Fisher R, Poole A, Thompson V (1979) Behavioural patterns within a general hospital psychiatric unit: an observational study. *Behav Res Ther* 17:317–332
- Shepherd G, Richardson R (1979) Organization and interaction in psychiatric day centres. *Psychol Med* 9:573–579
- Slade M, Phelan M, Thornicroft G (1996) The Camberwell Assessment of Need (CAN): comparison of assessments by staff and patients of the needs of the mentally ill. *Soc Psychiatry Psychiatr Epidemiol* 31:109–113
- Tugwell P (1979) A methodological perspective on process measures of the quality of medical care. *Clini Investig Med* 2:113–121
- Tyson G, Lambert G, Beattie L (1995) The quality of psychiatric nurses’ interactions with patients: an observational study. *Int J Nurs Stud* 32(1):49–58