REVIEW

A systematic review of the relationship between behavioral and psychological symptoms (BPSD) and caregiver well-being

Alexandra Feast,1 Esme Moniz-Cook,2 Charlotte Stoner,3 Georgina Charlesworth3 and Martin Orrell4

1Division of Psychiatry, University College London, London, UK
2Faculty of Health and Social Care, University of Hull, Hull, UK
3Research Department of Clinical, Educational, and Health Psychology, University College London, London, UK
4Institute of Mental Health, University of Nottingham, Nottingham, UK

ABSTRACT

Background: Behavioral and psychological symptoms in dementia (BPSD) are important predictors of institutionalization as well as caregiver burden and depression. Previous reviews have tended to group BPSD as one category with little focus on the role of the individual symptoms. This review investigates the role of the individual symptoms of BPSD in relation to the impact on different measures of family caregiver well-being.

Methods: Systematic review and meta-analysis of papers published in English between 1980 and December 2015 reporting which BPSD affect caregiver well-being. Paper quality was appraised using the Downs and Black Checklist (1998).

Results: Forty medium and high quality quantitative papers met the inclusion criteria, 16 were suitable to be included in a meta-analysis of mean distress scores. Depressive behaviors were the most distressing for caregivers followed by agitation/aggression and apathy. Euphoria was the least distressing. Correlation coefficients between mean total behavior scores and mean distress scores were pooled for four studies. Irritability, aberrant motor behavior and delusions were the most strongly correlated to distress, disinhibition was the least correlated.

Conclusions: The evidence is not conclusive as to whether some BPSD impact caregiver well-being more than others. Studies which validly examined BPSD individually were limited, and the included studies used numerous measures of BPSD and numerous measures of caregiver well-being. Future research may benefit from a consistent measure of BPSD, examining BPSD individually, and by examining the causal mechanisms by which BPSD impact well-being by including caregiver variables so that interventions can be designed to target BPSD more effectively.

Key words: behavioral and psychological symptoms of dementia, carer, dementia, neuropsychiatric symptoms, quality of life

Introduction

Behavioral and psychological symptoms in dementia (BPSD) can occur in up to 90% of people with dementia and have been associated with the breakdown of home care, increased psychiatric referral, misuse of medication, and increased healthcare costs (Cerejeira et al., 2012). BPSD are described as a major source of distress for the caregiver (Craig et al., 2005; Tan et al., 2005; Bandeira et al., 2007) and are composed of wide-ranging symptoms such as agitation, aggression, calling out repeatedly, sleep disturbance, wandering, and apathy. The majority of past research which has investigated the empirical relationship between BPSD and poor caregiver outcomes has used summative BPSD scores; therefore, it remains unclear as to whether specific individual symptoms impact caregivers differentially (Mohamed et al., 2010; Rocca et al., 2010).

The limited research which has distinguished between symptoms is mainly exploratory and does not control for the effect of multiple BPSD (Ballard et al., 1995; Gonzalez-Salvador et al., 1999; Neundorfer et al., 2001; Lim et al., 2008). However,
Ornstein et al. (2012) have advanced the field by examining how and why symptoms clusters impact caregiver depression. Nevertheless, the symptoms were clustered rather than individually represented and the systematic review did not employ a formal method of synthesis, but instead a descriptive analysis was given, along with tallying of symptoms.

It is often the case that BPSD are presented in a summative fashion due to the nature of the outcome measure and how the total score is calculated, and are consequently usually operationalized as a single construct of “behavior problem”; therefore, the data cannot be extracted. Research by Fauth and Gibbons (2014) highlighted the extensive inconsistencies in identifying BPSD that are the most problematic. These inconsistencies are partly due to whether “problematic” is defined and measured in terms of symptom prevalence, intensity, distress rating, or association with caregiver outcomes, such as depressive symptoms (Ornstein et al., 2012). A review by Black and Almeida (2004) has progressed our understanding of the association between BPSD as a whole, and the burden of care, and caregiver depression by performing a meta-analysis. However, multiple types of behavior measures, burden measures and measures of depression were pooled in the same meta-analysis. In future meta-analyses, it is important to combine the same outcome measures of a particular construct such as “distress” in order to ensure the statistical analysis is valid rather than combining multiple types of measures. This is addressed in the current review.

As recommended by Ornstein et al. (2012), it is necessary to design future systematic reviews which capture the complexity of unique behaviors in order to further our understanding of their differential impact upon caregivers. The aim of this study was to respond to Ornstein’s recommendation and to systematically review the relationship between individual BPSD and caregiver well-being in order to identify which BPSD are the most problematic for caregivers. This was achieved by distinguishing between psychological constructs of caregiver outcome (distress, burden, strain, stress) within each meta-analysis, and by distinguishing between how BPSD are measured (prevalence, frequency, and total score) within each meta-analysis.

Methods
Identification and selection of studies
A search strategy combining medical subject headings and text words relating to dementia, BPSD, mental disorders and behaviors, behaviors, elder care, significant others, carers, family, daughters, aged, carer burden, carer strain, carer distress, and cost of illness was devised and adapted for five electronic databases. Searches of studies published between 1980 and December 2015 were carried out through PsycINFO, CINAHL EBSCO (Cumulative Index to Nursing and Allied Health), MEDLINE, EMBASE, and Web of Science (including Science Citation Index-Expanded (SCI-EXPANDED), Conference Proceedings Citation Index-Science (CPCI-S), Social Science Citation Index (SSCI), Conference Proceedings Citation Index-Social Science & Humanities (CPCI-SSH), and reference lists of included studies. Studies were restricted to peer-reviewed original papers published in English and included participants who were family caregivers of someone with dementia, used a measure of caregiver response to BPSD and a measure of BPDS, and reported data for both measures. Reviews, conference proceedings, conference abstracts, theses, supplements, reports, letters, and non-peer reviewed papers were excluded.

Quality appraisal
Two reviewers (AF and CS) assessed the included papers for methodological quality, and disagreements were resolved by discussion. The 27-item Downs and Black Checklist (1998) was used to assess quality. Quality thresholds were those used in a previous systematic review (Crellas et al., 2014; ≥75% for high quality, ≥50% to 74% for medium quality). Agreement was measured using Cohen’s κ for closeness of scores.

Analysis
Summary statistics for each comparable caregiver outcome associated with each behavior (means, standard deviation, and sample size) were extracted for each study and were used to calculate the pooled weighted mean and standard deviation for that outcome. The following formulae were used:

Mean for combined studies: \( \frac{\sum n_i m_i}{\sum n_i} \), where \( n_i \) is the sample size for study 1 and \( m_i \) is the mean for study 1 (Price, 1970). The standard deviation of the combined studies: \( \sqrt{\left( \sum_n (n_i - 1) \left( \frac{(sd_i)^2}{\sum n_i} \right) - T \right)} \), where \( T \) is the number of studies (all summations from 1 to \( T \) (Cohen, 1988). Standard deviations were converted to standard errors: \( SE = SD/\sqrt{N} \).

MedCalc software (Version 15: Ostend, Belgium) was used to perform a meta-analysis of comparable studies where possible in order to produce a pooled effect. Studies with smaller standard errors and larger sample sizes were given more weight in the calculation. The \( I^2 \) statistic was calculated to denote the percentage of observed total variation across studies that is due
Individual BPSD and caregiver well-being

1763

Records identified through database searching (n = 12,244)
MEDLINE Ovid = 3,618
EMBASE Ovid = 576
PsycINFO Ovid = 1,502
CINAHL EBSCO = 1,056
OpenGrey = 30
Web of Science = 5,454

Records after duplicates removed (n = 9,485)

Records excluded following the screening of titles and abstracts (n = 8,892)

Full-text articles excluded n = 553
(No data on the association between BPSD and distress n = 425)
(Not English n = 32)
(No dementia n = 20)
(Not a family carer n = 22)
(Review n = 35)
(Report n = 1)
(Residential care n = 17)
(Thesis n = 1)

Hand searched reference lists (n = 8)

Studies included (n = 32)

Total studies included (n = 40)

Figure 1. Study selection process.

to heterogeneity. A p value of >0.10 determines if heterogeneity is present, consequently the summary effect from the random-effects model rather than the fixed-effects model is utilized. In cases where data were not amenable to meta-analysis (either not comparable, or insufficient data), these studies were compared descriptively and tabulated.

Results

A total of 12,244 references were identified (Figure 1), of which 11,651 were excluded by screening the title, abstract, and the removal of duplicates. Of the remaining 593 references, full texts were sought, and 40 met the inclusion criteria regardless of quality. Reasons for exclusion can be seen in Figure 1.

The 40 included studies were quantitative, mainly cross-sectional with two longitudinal studies (Berger et al., 2005; Johnson et al., 2013), 21 studies were graded as high quality, and 19 as medium quality. A brief summary of the demographic data of the studies is provided in Table S1, along with a description of the measures used, quality score, and the main findings. There was little consistency in the sample populations across studies. The included studies consisted of 10,978 caregivers in total, of these 67% were female and had a mean age of 60 years. Only five studies reported ethnicity. Of those which reported dementia stage, nine studies included participants with mild dementia, 20 with moderate, and one with severe dementia.

Quality appraisal of included studies

Level of agreement between the two independent reviewers was $\kappa = 0.72$, $p = 0.002$, 95% CI (0.73, 0.78) indicating substantial agreement (Landis and Koch, 1977). The studies scoring the lowest on the checklist were Haley et al. (1987), Hishikawa et al. (2012), and Savorani et al. (1998). All studies lost points on questions covering external validity,
validity, and reliability of outcome measures and the reporting of statistics.

**Meta-analysis of comparable studies**

Sixteen of the 40 studies were suitable to be included in a meta-analysis of mean distress scores (see Table S2, available as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG). All of these studies reported data on the mean and standard deviation of distress associated with the mean total score on the NPI for individual behaviors. Each adapted version of the NPI was reported as valid and reliable and all were administered by a researcher or clinician. Five studies utilized the ten-item version of the NPI; therefore, data reporting distress associated with sleep and appetite/eating change were unavailable for these studies. Participants were recruited from clinic-based and population-based studies, had a mix of dementia participants and studies only including those with a diagnosis of Alzheimer’s disease, Parkinson’s disease dementia, and frontotemporal dementia (see Table S1, available as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG). In Table S3 (available as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG), the mean distress reported for the total NPI symptom can be seen for each of the 16 studies, along with the pooled standard deviations for each behavior across the studies. There is considerable variation between the studies for each of the symptoms, this can be seen in Figure 2, in particular, there is substantial variation in the mean distress reported for the total score of irritable behavior: Baiyewu et al. (2003) reported 0.6 (between “not at all distressing” and “minimally distressing”) and Balieiro et al. (2010) reported 3.6 (between “moderately” and “severely” distressing). However, it is possible to conclude that euphoria was the least distressing across the studies.

As seen in Figure 3, the most distressing BPSD across the 16 studies was depression. Agitation/aggression was the second most distressing, followed by apathy. Euphoria was the least distressing and had the smallest standard error. Delusional and euphoric behaviors were the only two behaviors with a standard deviation exceeding the mean indicating high variance between studies for these behaviors.

Four studies reported correlation data between mean NPI total scores and mean NPI distress (Matsumoto et al., 2007; Balieiro et al., 2010; Davidsdottir et al., 2012; Huang et al., 2012), all of these studies were involved in the previous meta-analysis other than Davidsdottir et al. (2012). Three of these studies reported data on all 12 behaviors. The sample size and the correlation coefficients were entered into MedCalc in order to produce a pooled effect size for each behavior. The pooled effect sizes within the fixed-effect model, random-effects model, and heterogeneity statistics are reported in Table 1. Significant heterogeneity was identified between studies for all behaviors other than apathy and sleep. In contrast to the mean distress data, the meta-analysis of the correlation data identified that irritable behavior was the most strongly associated with caregiver distress followed by aberrant motor behavior and then delusions. Disinhibition had the weakest association with distress. Depression was seen to be one of the behaviors which was more strongly associated/correlated with distress across both meta-analyses. Appetite and eating-related behavior was one of the behaviors which had a weaker association/correlation with distress across both meta-analyses.

**Non-comparable studies**

Twenty of the 40 included studies were not suitable for meta-analysis. These studies were not comparable in terms of BPSD measure and the associated unit of measurement, caregiver well-being measure, and the associated unit of measurement. It was therefore only possible to compare these 20 studies descriptively due to their heterogeneity.

**Prevalence of BPSD and the impact on caregiver well-being**

Mean prevalence and mean distress were reported on the NPI by Haley et al. (2008) and de Vugt et al. (2006). Although these two studies both measure prevalence of BPSD, methods for measuring caregiver well-being differed (see Table S1).

**Frequency of BPSD and caregiver well-being**

Memory-related behaviors were reported as most stressful and problematic across three studies (Nygaard, 1998; Quayhagen and Quayhagen, 1998; Muangpaisan et al., 2010). However, Robinson et al. (2001) and Fauth and Gibbons (2014) both found the frequency of memory-related behaviors was associated with the least amount of reactivity. Disruptive behaviors were associated with the most reactivity, followed by depressive behaviors (Robinson et al., 2001; Fauth and Gibbons, 2014). In all of the studies which measured hallucinations caregivers reported this behavior...
as less problematic, burdensome and distressing compared to many of the other behaviors. Verbal aggression across three of the five studies was reported as problematic and burdensome. Across two of the five studies which reported sexual disinhibition, this behavior was reported as the least problematic. Reports of wandering across four of the studies were not consistent, since some of the studies reported this behavior as much more problematic and burdensome than others.

Across four of the five studies which reported incontinence, this was consistently reported as one of the more burdensome, stressful and problematic behaviors. Physical aggression was only measured in three studies and only one study reported this
behavior as a common problem. Both studies which reported irritability identified this behavior to be one of the most problematic and burdensome.

**Correlational data between BPSD and caregiver well-being**

Sleeping- and eating-related behaviors were found to be significantly correlated to burden and strain; however, they had a stronger association with burden (Donaldson et al., 1998). With regard to adult-child caregivers, four (delusions, agitation, disinhibition, and appetite/eating change) of the eight behaviors reported by Conde-Sala et al. (2010) as being significantly associated with burden were also found to be significantly associated with burden by Allegri et al. (2006). Apathy and irritability were reported to be significantly associated with burden in the study conducted by Conde-Sala et al. (2010), but no significant relationship was reported by Allegri et al. (2006). Spousal caregivers only found appetite/eating change burdensome (Conde-Sala et al., 2010). Apathy had a moderate significant association with burden in spousal and adult-child caregivers (Conde-Sala et al., 2010), but this association was non-significant in the data reported by Allegri et al. (2006). Delusional and irritable behaviors were also reported as non-significantly related to burden (Allegri et al., 2006) but were significantly related to burden by Conde-Sala et al. (2010). In contrast to the previous studies reporting the weak association between euphoria and distress, Fuh et al. (2001) reported a significant correlation between euphoria and burden. No other behaviors on the NPI had a significant correlation with burden (Fuh et al., 2001).

In agreement with the data concerning at the association between the frequency of the behavior within each subscale on the RMBPC and reactivity, disruptive behaviors had the largest correlation of the three subscales with burden (Huang et al., 2012). Using the BEHAVE-AD (Reisberg et al., 1982) and the NOSGER (Spiegel et al., 1991) mood, social behaviors, and disturbing behaviors were found to have a significant moderate association with burden but no significant correlations with depression in caregivers. Chiu et al. (2014) noted that all of the behaviors on the NPI significantly predicted sleep disturbance in caregivers other than anxiety and aberrant motor behavior.
_valid conclusions were drawn by only synthesizing caregiver well-being are measured, which ensured valid conclusions were drawn by only synthesizing inferences drawn. The results are discussed in the context of previous research.

### Discussion

To our knowledge, this is the first review to control for the effect of multiple BPSD on caregiver well-being by discriminating between individual BPSD, and also discriminating between caregiver outcomes of well-being. Additional advantages of this review were the use of a formal method of synthesis and the decision to distinguish between how BPSD and caregiver well-being are measured, which ensured valid conclusions were drawn by only synthesizing comparably data. Only two meta-analyses were possible, due to limited homogeneous studies. Categories in which the relationship between BPSD and well-being were qualitatively examined included minimal studies displaying considerable variation; therefore, valid conclusions could not be drawn. The results are discussed in the context of previous research.

### Meta-analysis of mean distress scores

As seen in Figures 2 and 3, even within one measure and one criterion (distress associated with total score on the NPI) variability in distress ratings for each symptom can be seen between the 16 studies. This was also evident in research conducted by Fauth and Gibbons (2014), in which a wide variation in the prevalence of NPI symptoms was seen between studies. However, data regarding distress associated with euphoria and delusions demonstrated high levels of variance in terms of the ratio between the mean and the standard deviation resulting in considerable variability between scores. Therefore, the ratio of the multivariate distress ratings compared with the standard deviation associated with the other ten behaviors signifies that there is low variance between the 16 studies. It is unclear at present why there is increased variability for euphoric and delusional behaviors. In terms of total score, depression followed by agitation/aggression behavior were stronger predictors of distress as measured by the NPI. Apathy was a significant predictor across two of the three studies. Delusions and anxiety were the only other comparable behaviors across two of the three studies and were both significant predictors of distress. Covinsky et al. (2003) did not use a formal behavior scale but identified that anger and aggressive behaviors were stronger predictors of caregiver depression, followed by dangerous behaviors to self and others, and then behaviors which wake the caregiver up at night.

### Predictive data between BPSD and well-being

Conde-Sala et al. (2010), Chappell and Penning (1996), Davis and Tremont (2007), and Slachevsky et al. (2013) used the same burden measure, but used different measures of BPSD. Apathy significantly predicted burden across three studies, although it was less burdensome for adult-child caregivers and did not predict apathy in the study by Davis and Tremont (2007). Three studies (Mourik et al., 2004; Khoo et al., 2013; Lee et al., 2013) reported data identifying which behaviors predicted distress as measured by the NPI. Apathy was a significant predictor across two of the three studies. Delusions and anxiety were the only other comparable behaviors across two of the three studies and were both significant predictors of distress.

### Table 1. Meta-analysis of mean distress scores correlated with behaviors using NPI total scores

<table>
<thead>
<tr>
<th>BEHAVIOR</th>
<th>N</th>
<th>POOLED CORRELATION COEFFICIENT</th>
<th>TEST FOR HETEROGENEITY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FIXED-EFFECTS MODEL (95% CI)</td>
<td>RANDOM-EFFECTS MODEL (95% CI)</td>
</tr>
<tr>
<td>Delusions</td>
<td>243</td>
<td>0.557 (0.462 to 0.640)</td>
<td>0.567 (0.412 to 0.690)</td>
</tr>
<tr>
<td>Hallucinations</td>
<td>243</td>
<td>0.502 (0.399 to 0.532)</td>
<td>0.520 (0.198 to 0.791)</td>
</tr>
<tr>
<td>Agitation/Aggression</td>
<td>243</td>
<td>0.540 (0.443 to 0.625)</td>
<td>0.537 (0.299 to 0.712)</td>
</tr>
<tr>
<td>Depression</td>
<td>243</td>
<td>0.528 (0.429 to 0.615)</td>
<td>0.564 (0.360–0.717)</td>
</tr>
<tr>
<td>Anxiety</td>
<td>243</td>
<td>0.482 (0.377 to 0.575)</td>
<td>0.481 (0.236 to 0.669)</td>
</tr>
<tr>
<td>Euphoria</td>
<td>243</td>
<td>0.540 (0.443 to 0.625)</td>
<td>0.480 (0.147 to 0.715)</td>
</tr>
<tr>
<td>Apathy</td>
<td>243</td>
<td>0.456 (0.351 to 0.554)</td>
<td>0.459 (0.351 to 0.540)</td>
</tr>
<tr>
<td>Disinhibition</td>
<td>243</td>
<td>0.357 (0.240 to 0.464)</td>
<td>0.371 (0.162 to 0.548)</td>
</tr>
<tr>
<td>Irritability</td>
<td>243</td>
<td>0.668 (0.590 to 0.733)</td>
<td>0.691 (0.486 to 0.823)</td>
</tr>
<tr>
<td>Aberrant motor behavior</td>
<td>243</td>
<td>0.606 (0.518 to 0.681)</td>
<td>0.603 (0.437 to 0.730)</td>
</tr>
<tr>
<td>Sleep</td>
<td>176</td>
<td>0.438 (0.307 to 0.552)</td>
<td>0.436 (0.301 to 0.554)</td>
</tr>
<tr>
<td>Appetite and eating</td>
<td>176</td>
<td>0.405 (0.271 to 0.523)</td>
<td>0.442 (0.200 to 0.633)</td>
</tr>
</tbody>
</table>
and distress and found similar results; however, even with identical analysis there were noteworthy discrepancies, such as delusions being the most distressing but only the fourth most distressing in the current review. Consistent with the current meta-analysis of mean data, apathy, and depression-related behaviors have been reported as distressing in numerous studies (Teri, 1997; deVugt et al., 2003; Aarsland et al., 2007; Kosberg et al., 2007). Similarly, depression in the person with dementia is associated with depression in the caregiver (Waite et al., 2004).

Meta-analysis of correlation data

As seen in Table 1, there were high levels of heterogeneity between the studies included in the second meta-analysis investigating which behaviors were associated with distress. Only apathy- and sleep-related behaviors displayed homogeneity between studies. Furthermore, behaviors which had stronger correlations with distress (Table 1) did not have strong associations in the previous meta-analysis (Figure 3) regarding distress associated with behaviors. At present, it is unclear why this discrepancy exists, and also why distress correlated with apathy- and sleep-related behaviors is more homogenous between studies than the other behaviors. Since only a few studies contributed to this meta-analysis, more research is needed to investigate the pooled effects. Nevertheless, a meta-analysis of this nature has not been performed previously and the random-effects model was used to report a more conservative estimate for the heterogeneous pooled effects in order to reduce bias.

Apathy, and irritable behaviors were not consistently significantly correlated to burden across multiple studies, whereas delusions, agitation, disinhibition, and appetite/eating change were; these associations were also reported in other studies (Rymer et al., 2002; Rocca et al., 2010). In contrast, a review by Ornstein and Gaugler (2012) reported that anger/aggression and depression were the most frequently cited BPSD associated with caregiver burden; depression was also linked to caregiver burden in a review by Black and Almeida (2004). In addition to burden, delusional behavior was consistently associated with distress across studies, with a significant correlation also being seen with stress (Kosmala and Kłoszewska 2004). Berger et al. (2005) noted that behaviors which are significantly correlated with burden are not correlated with depression. This demonstrates the importance of including multiple measures of well-being since the same behaviors affect caregivers differently, which has implications for future interventions.

Studies unamenable to quantitative synthesis

Only two studies reported the prevalence of BPSD and the impact on caregiver well-being; due to differences between studies, comparisons were not possible. Although five studies reported the frequency of BPSD, each measure of caregiver well-being was different. Memory-related behaviors, verbal aggression, incontinence, and irritability were reported to affect well-being across multiple studies: These findings were also seen in previous research (Bruce and Paterson, 2000; Farran et al., 2004). However, Fauth and Gibbons (2014) and Robinson et al. (2001) did not find that memory-related behaviors were distressing despite being the most frequent type of behavior on the RMBPC. Hallucinations, sexual disinhibition, and physical aggression were consistently reported to affect well-being less than the other behaviors. Although this finding is reflected in the majority of studies, Haley et al. (2008) and Arango-Lasprilla et al. (2010) reported contradictory results.

It is noteworthy that aggression/agitation, when measured in terms of total score, was reported as the second most distressing across 16 of the studies, but when measured in terms of frequency only one of the three studies reported physical aggression as a common problem. However, the frequency of verbal aggression consistently affected well-being. It is not possible to provide an explanation for this discrepancy, but it may be because of the distinction between how aggression/agitation is measured (i.e. total score or frequency), or that verbal and physical aggression have been differentiated and therefore affect caregiver well-being differently.

Predictive data were limited and heterogeneous in the present review; however, it was possible to conclude that apathy predicted burden and distress across studies. Anger, restlessness, combativeness, dangerous to others, and embarrassing behavior were found to be strong predictors of burden in a study by Gaugler et al. (2011). Only one study reported the predictors of caregiver depression.

Potential explanations for variability in caregiver well-being

Difference in study quality does not appear to be an explanation for the variation in distress rating across studies, since even between two medium quality studies, there was considerable variation (Balieiro et al., 2010, displaying one of the highest mean distress scores; Baiyewu et al., 2003, displaying one of the lowest). Two studies consistently had the highest mean distress rating across each of the behavior domains: Balieiro et al. (2010) and Tan et al. (2005). Both of these studies were clinic-based, higher distress ratings would be expected as those
experiencing more BPSD might be more likely to attend a clinic, as, indeed, were the majority of studies included in the meta-analysis. It may be possible that dementia severity is an explanation, since participants in the studies with the highest distress scores (Tan et al., 2005; Balieiro et al., 2010) had moderate and severe dementia respectively, whereas Baiyewu et al. (2003), who consistently reported the lowest distress scores across each of the behavior domains, included participants with mild dementia. However, both Aarsland et al. (2007) and de Vugt et al. (2006) reported lower mean distress scores in comparison with the other included studies, and they both included participants with moderate dementia. Furthermore, Davidsdottir et al. (2012) found that there was no significant difference between caregiver distress scores depending on dementia severity.

Baiyewu et al. (2003) had a much smaller sample of spousal caregivers in comparison to Tan et al. (2005), which might account for the lower distress scores. This is in agreement with Almberg et al. (1997) who reported that spouses experience greater levels of burden. However, three studies which had consistently higher mean distress scores across all behaviors (Tan et al., 2005; Balieiro et al., 2010; Huang et al., 2012), were mainly comprised of adult-child caregivers. When comparing a study solely comprised of spousal caregivers (deVugt et al., 2006) with a study (Baiyewu et al., 2003) with minimal spousal caregivers (2.5%), it was possible to see that spousal caregivers were substantially more distressed with depressive, apathetic, and anxiety-related behaviors. Although distress ratings can be differentiated between behavior types for spousal and non-spousal caregivers, both Baiyewu et al. (2003) and de Vugt et al. (2006) had two of the lowest distress scores across all behaviors in comparison to the other studies in the meta-analysis.

The explanation may also be cultural since Baiyewu et al. (2003) and Baiyewu et al. (2012) (with similarly low distress scores across behaviors) were the only Africa-based studies in the meta-analysis. Baiyewu et al. (2003) and Baiyewu et al. (2012) were also both community-based studies which may also be a reason for the reduced distress ratings. There is every reason to believe that BPSD are present across cultures (Shaji et al., 2009); however, the low levels of public awareness of dementia have many implications (Prince, 2000). One of these implications may be the misinterpretation of BPSD as deliberate, consequently affecting the well-being of the carer (Shaji et al., 2009). It is beyond the scope of this paper to determine why there is considerable variation between studies reporting caregiver distress for identical behaviors using identical measures. However, it is likely that it is the caregiver characteristics rather than the characteristics of the person with dementia (such as dementia severity) which may underlie the explanation as to why variation between studies exists.

Limitations

The data available for this review allowed for only two meta-analyses to be conducted. There were relatively few studies which assessed the relationship between individual BPSD and caregiver well-being and which met the inclusion/exclusion criteria for the review. In an attempt to be systematic based on prior recommendation, an effort was made to distinguish between how “problematic” symptoms are measured. Consequently, studies were categorized into whether BPSD were measured in terms of frequency or prevalence, and also categorized in terms of the statistical analysis employed. Other than data regarding mean distress associated with NPI total score, only a small amount of studies were included in each category and unfortunately, it was often the case that measures of caregiver well-being were dissimilar. From the 40 included studies, there were 12 different measures of behavior and 17 different measures of well-being, making comparisons difficult which may have contributed to the inconsistent findings.

The majority of included studies recruited participants from clinics; therefore, caregivers included in the review were not representative of the population. In addition, the samples across the included studies were heterogeneous in a number of factors; therefore, caution should be taken when interpreting the results. Although beyond the scope of this review, study characteristics such as culture, gender, and relationship direction were not examined through subgroup analyses. Due to the limited number of studies, it was not possible to impose a quality criteria; however, all included studies were either of high or medium quality as assessed by the Downs and Back Checklist (1998). Furthermore, all of the included studies relied on caregiver reports of BPSD, which may themselves be affected by caregiver well-being; unfortunately, it is not possible to determine if this was the case. Nonetheless, two valid meta-analyses were completed and a number of comments can be made about the relationship between BPSD and caregiver well-being.

Implications and recommendations

The vast majority of studies addressing the relationship between BPSD and caregiver well-being
did not examine individual BPSD. Consequently, this comparatively limited the amount of studies included in the present review. Therefore, if we are to understand this relationship, more studies need to examine individual BPSD in their research. In addition, it is more clinically relevant to examine which BPSD affect well-being rather than BPSD as a whole, as this is essential in determining where clinical staff should allocate resources most effectively. Nevertheless, there is preliminary data suggesting it is mood-related behaviors which are associated with increased levels of distress; therefore, interventions for the management of BPSD should target these behaviors.

The lack of consistency in the measurement of BPSD (neuropsychiatric or memory problems, frequency, or prevalence), and the measurement of caregiver well-being (the use of interchangeable terminology such as stress, strain, and burden meaning the same construct) limits our ability to reach a consensus as to which BPSD result in the most negative outcomes for caregivers. For example, Donaldson et al. (1998) used Gilleard’s Strain Scale (1984) to measure burden and the General Health Questionnaire (Goldberg and Hillier, 1979) to measure distress. Furthermore, four studies (see Table S1) did not have formal scales to measure the impact of BPSD on well-being, and used terms such as “burdensome,” “serious,” and “disturbing.” This problem impacts upon the validity and reliability of the original studies and consequently the review. However, this mainly occurred in the earlier studies. Future research should also include more than one measure of well-being in order to identify if individual BPSD differentially affect caregiver burden compared to caregiver distress as identified in the current review. Furthermore, future research should identify why there is reduced variability in distress ratings for certain behaviors compared to others, and why certain behaviors are more distressing depending on which type of statistical analysis is used.

Although there are potential methodological reasons (limited studies, inconsistent measures) for why there is substantial variability when identifying which BPSD affect certain types of caregiver well-being, the most, there is growing evidence that factors associated with the person with dementia (e.g. BPSD frequency, cognition) are only part of the story when explaining a caregiver’s response to BPSD. Caregiver characteristics have been shown to account for twice the variance of person with dementia characteristics when reporting neuropsychiatric symptoms (Sink et al., 2006), and the presence of BPSD was not found to be a significant predictor of burden, whereas caregiver characteristics such as confidence, neuroticism, and relationship quality were significant predictors (Campbell et al., 2008). Therefore, as recommended by Ornstein and Gaugler (2012), future studies should focus on identifying pathways for the association between individual BPSD, caregiver variables and caregiver outcomes. More research is needed to understand if and why certain BPSD affect certain caregiver outcomes, and also how much of this is attributable to caregiver variability in response to BPSD rather than the presence of BPSD itself. More longitudinal studies would be useful to examine how changes in individual BPSD are associated with caregiver well-being over time. A review by Black and Almeida (2004) examined the limited longitudinal studies available and reported mixed findings between burden and BPSD, and found no association between BPSD and distress. Although BPSD increased significantly over a two-year period, caregiver burden remained stable, and depression was found to decrease (Berger et al., 2005). Future research needs to examine caregiver factors over time (well-being, role captivity, etc.) along with changes in individual BPSD in order to have a more valid understanding of this relationship.

Caution should be exercised when creating a measure of BPSD, since Jackson et al. (2014) found that the Revised Memory and Behavior Problem Checklist (Teri et al., 1992) might be more suitable in explaining caregiver depression than the NPI. It is important that the measure of BPSD is not biased toward explaining a more unique association between BPSD and a measure of caregiver well-being such as depression. As stated by Gitlin and Rose (2014), item selection for measures of BPSD is often based upon specific disciplines (e.g. neurology, psychiatry, social sciences) and thus implicitly represents different conceptualizations of behavior. A shared framework for defining behaviors and their underlying causes must be created in order to move on from the implicit assumption of most measures that behaviors are a direct cause of neuropathology, and to develop measures which also address the contextual and caregiver characteristics associated with the behavior.

**Conclusions**

This study has improved our understanding of the impact of individual BPSD on caregiver well-being, in particular caregiver distress. The relationship between individual BPSD and caregiver well-being was inconsistent and varied according to behavioral symptom. Due to limited studies and heterogeneous data, the results should be interpreted with caution,
but advances have been made with respect to which behaviors are associated with distress. The inconsistencies may be partly explained by caregiver variables which were not taken into account in the review. Based on this review and previous research, it is not possible to fully understand the relationship between BPSTD and caregiver well-being without examining the influence of caregiver variables such as caregiver strategies, acceptance, gender, relationship with person with dementia, and confidence. We recommend that future studies are designed to examine individual BPSTD rather than BPSTD as a whole, use consistent measures for BPSTD and well-being, use consistent terminology for psychological constructs (e.g. burden), and examine the causal mechanisms by which individual BPSTD impact caregiver well-being by collecting data on a range of caregiver factors. These approaches will clarify whether it is important for clinicians to focus on the frequency or prevalence of certain BPSTD, in addition to targeting the presence of certain caregiver characteristics and addressing unmet need.

Conflict of interest
None.

Description of authors’ roles
A. Feast was responsible for developing the review objectives, writing the search strategies, conducting the searches, designing and completing study eligibility and data extraction forms, undertaking the quality appraisals, conducting the meta-analysis, and writing the paper. M. Orrell was responsible for developing the review objectives, advising on the review methodology, meta-analysis, and assisting in writing the paper. G. Charlesworth was responsible for undertaking the quality appraisals and assisting in writing the paper. C. Stoner was responsible for undertaking the quality appraisals and assisting in writing the paper. E. Moniz-Cook was responsible for developing the review objectives, advising on the review methodology, meta-analysis, and assisting in writing the paper.

Acknowledgments
We would like to thank Dr Zoë Hoare for her statistical advice. This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research (RP-PG-0606-1067 and RP-PG-060-1083). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Supplementary Material
To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1041610216000922.

References


**Gonzalez-Salvador, M. T., Arango, C., Lyketsos, C. G. and Barbá, A. C.** (1999). The stress and psychological


