Delirium Burden in Patients and Family Caregivers: Development and Testing of New Instruments

Delirium Burden in Patients and Family Caregivers:

Development and Testing of New Instruments

Annie M. Racine[a,b,c]; Madeline D'Aquila[a]; Eva M. Schmitt[a]; Jacqueline Gallagher[e]; Edward R.

Marcantonio[a,e]; *Richard N. Jones[e]; *Sharon K. Inouye[a,d]; *Dena Schulman-Green[d]

for the BASIL Study Group

*Co-senior authors

[a]Aging Brain Center, Institute for Aging Research, Hebrew SeniorLife, [b]Harvard Medical School, Boston,

MA; [c]Frontotemporal Disorders Unit, Massachusetts General Hospital, Boston, MA; [d]Yale School of

Nursing, West Haven, CT; [e] Departments of Medicine, Beth Israel Deaconess Medical Center, Harvard

Medical School, Boston, MA; [f]William F. Connell School of Nursing, Boston College, Boston, MA;

[g]Brown University Warren Alpert Medical School, Providence RI

Corresponding author: Sharon K. Inouye, M.D., MPH; Professor of Medicine, Harvard Medical School

Division of Gerontology, Beth Israel Deaconess Medical Center; Milton and Shirley F. Levy Family

Chair; Director, Aging Brain Center, Institute for Aging Research, Hebrew SeniorLife; Address: 1200

Centre Street, Boston, MA 0213; Phone: 1-617-971-5390; Fax: 1-617-971-5339; Email (not for

publication): sharoninouye@hsl.harvard.edu; Email for publication: agingbraincenter@hsl.harvard.edu

Author email addresses: AnnieRacine@hsl.harvard.edu, MadelineDAquila@hsl.harvard.edu,

EvaSchmitt@hsl.harvard.edu, jgallag2@bidmc.harvard.edu, emarcant@bidmc.harvard.edu,

richard_jones@brown.edu, SharonInouye@hsl.harvard.edu, dena.schulman-green@yale.edu

Word count (abstract, manuscript, references): 4,937

Field Team: Tatiana Abrantes, BS (HSL); Brett Armstrong, MPH (BIDMC); Sylvia Bertrand, BS (HSL); Angelee Butters, MA (BIDMC); Madeline D'Aquila, BS (HSL); Jacqueline Gallagher, MS (BIDMC); Jennifer Kettell, BS (HSL); Jacqueline Nee, BA (HSL); Katelyn Parisi, BA, (HSL); Margaret Vella, BS (HSL); Guoquan Xu, MD, PhD (HSL); Lauren Weiner, MA (BIDMC).

Data Management and Statistical Analysis Team: Yun Gou, MA (HSL); Douglas Tommet, MPH (Brown University).

Expert Review Panel: Charles H. Brown, M.D. (Johns Hopkins); Sevdenur Cizginer, M.D. (Brown University); Diane Clark, PT, DScPT, MBA (University of Alabama); Joseph H. Flaherty, MD (St. Louis University); Anne Gleason, B.S. (HSL); Ann M. Kolanowski, Ph.D., RN (Penn State); Karen J. Neufeld, MD, MPH (Johns Hopkins University); Margaret G. O'Connor, PhD (BIDMC);Margaret A. Pisani, M.D., MPH (Yale); Thomas Robinson, M.D. (University of Colorado); Joe Verghese, M.B., B.S. (Albert Einstein); Heidi Wald, M.D., MPH (University of Colorado);Sharon M. Gordon, Psy.D. (Vanderbilt)

Abstract

Background and Objectives: Delirium creates distinct emotional distress in patients and family caregivers, yet there are limited tools to assess the experience. Our objective was to develop separate patient and family caregiver delirium burden instruments and to test their content and construct validity.

Research Design and Methods: 247 patients and 213 family caregivers were selected from an ongoing prospective cohort of medical-surgical admissions aged ≥70 years old. New patient and family caregiver delirium burden instruments were developed and used to measure the subjective experiences of in-hospital delirium. Delirium and delirium severity were measured by the Confusion Assessment Method (CAM) and CAM-Severity (long form).

Results: Both Delirium Burden (DEL-B) instruments consist of 8 questions and are measured on a 0-40 point scale. Final questions had good clarity and relevancy, as rated by the expert panel, and good internal consistency (Cronbach's alpha=.82-.86). In the cohort validation, Patient DEL-B (DEL-B-P) was 5.1 points higher and Family Caregiver DEL-B (DEL-B-C) was 5.8 points higher, on average, for patients who developed delirium compared to those who did not (p<.001). Test-retest reliability of DEL-B-C at baseline and one-month was strong (correlation=0.73). Delirium severity was mildly-moderately correlated with DEL-B-P (correlation=.34) and DEL-B-C (correlation=.26), suggesting contribution of other factors.

Discussion and Implications: We created instruments to reliably measure and evaluate the burden of delirium for patients and their family caregivers. While additional validation is indicated, these instruments provide a key first step towards measuring and improving the subjective experience of delirium for patients and their families.

Key words: delirium burden; distress, instrument development, family caregiver, caregiver burden, patient burden

Delirium Burden in Patients and Family Caregivers: Development and Testing of New Instruments

Introduction

Delirium, an acute decline of global cognitive functioning, is a common, serious, and costly complication of hospitalization for older adults, with incidence rates of 29-64% (Inouye, Westendorp, & Saczynski, 2014; Marcantonio, 2018). While prior research in delirium has focused on medical aspects of prevention and treatment, recent studies have underscored its human toll particularly on patients and family caregivers (Finucane, Lugton, Kennedy, & Spiller, 2017; Fuller, 2016; Morandi et al., 2012; Partridge, Martin, Harari, & Dhesi, 2013). Delirium burden, defined here as the subjective experience of delirium for patients and family members, includes awareness of delirium symptoms, situational stress, and emotional response. Delirium is often distressing to patients and their family caregivers, and delirium burden may continue to affect emotional, psychological, and physical well-being long after the delirium episode resolves. A better understanding of delirium burden may help to quantify these important aspects of the lived delirium experience and to develop better strategies to support patients and their family caregivers during and after a delirium episode.

To date, there are limited tools to assess delirium burden. Family caregiver burden in delirium has been studied using the Caregiver Burden Scale (Zarit, Reever, & Bach-Peterson, 1980), the Caregiver Burden Inventory (Novak & Guest, 1989), and the Older Persons and Informal Caregivers Survey Minimum DataSet (Lutomski et al., 2013); however, these tools are not delirium specific, and do not fully capture the burden of delirium (Breitbart, Gibson, & Tremblay, 2002; Bruera et al., 2009; Grover, Ghosh, & Ghormode, 2015; Morandi et al., 2012) associated with its acuity, unpredictability, and unique symptoms, such as inattention, visual hallucinations, and delusions. While caregiver burden has been examined extensively in dementia (Chiao, Wu, & Hsiao, 2015), these measures emphasize long-term emotional, physical, and economic strains, which are common in dementia care, but do not translate well to the acute situation of delirium. The Delirium Experience Questionnaire (DEQ) (Breitbart et al., 2002),

a brief instrument developed to assess delirium recall and distress in patients, caregivers, and nurses has

not yet been fully validated but represented an important advance.

Our objective was to develop two separate instruments to assess delirium burden among patients and

family caregivers that would be sensitive to the acuity and unique features of delirium, and that could be

applied in both clinical and research settings. We followed guidelines for instrument development,

including identifying content domains and their operational definitions, generating items, expert review of

items, and testing in a target population (DeVellis, 2016; Lynn, 1986). We hypothesized that delirium

burden would be greater for both patients and family caregivers in patients with delirium compared to no

delirium and would increase with increasing delirium severity; thus, we investigated convergent validity

with delirium incidence and severity.

Design and Methods

Content domains and operational definitions

Content domains were based on our previous qualitative work which identified three themes

common to patients' and family caregivers' delirium experience: symptom burden, emotional burden, and

situational burden (Schmitt et al., 2017). Symptom burden encompasses burden related to experiencing

(patient) or observing (family) symptoms of delirium, including disorientation, hallucinations, delusions,

impaired communication, memory problems, personality changes, and sleep disturbances. Emotional

burden encompasses emotions associated with delirium, such as anger, frustration, fear, guilt, and

helplessness. Situational burden is triggered by having to manage delirium and includes loss of control,

lack of support (e.g., lack of knowledge or resources to care for a delirious patient), safety concerns,

unpredictability, and feeling unprepared. Based on first-hand accounts from our qualitative interviews

supplemented by literature review (Schmitt et al., 2017), these content domains formed the basis for

designing delirium burden instruments for both groups.

Generating items and expert review

Our study team systematically generated a potential list of experts to represent relevant disciplines and content expertise, who are widely recognized as experts in the field. Eight agreed to participate who provided multidisciplinary perspectives (**Supplementary Table 1**) with extensive expertise in delirium, as well as relevant clinical and methodological expertise. The expert panel was varied in terms of discipline and expertise (e.g., anesthesiology, critical care, geriatrics, hospital medicine, internal medicine, neurology, nursing, occupational therapy, psychiatry, surgery); educational training (2 MD, 1 PhD, 1 PsyD, 2 BS, 1 RN, 2 MPH; several had multiple degrees and specialties); geographic location (5 northeast, 1 southeast, 2 southwest); and gender (63% female).

We generated an initial set of 17 items for patients and 23 for family caregivers that reflected the three content domains (**Tables 2-3**). We circulated the initial set of items for each instrument to the expert panel along with a content validity rating form **(see Supplementary Figures 1-2)**. Experts rated each item on perceived relevance to delirium burden and on clarity using a 1-4 scale (not relevant/clear - very relevant/clear). Some items were repeated to present various wording options. We asked experts to select the five best items ("top 5"), to comment in an open-ended fashion on individual items or on the instrument as a whole, and to suggest additional items.

Finalization of items

We averaged and evaluated experts' rankings of relevance and clarity for each item. We did not consider items rated as less than moderately relevant (on average) for inclusion in the final instrument. We adjudicated the remaining items based on "top 5" ratings, qualitative feedback, and face validity, and modified final items based on expert suggestions. We used Cronbach's alpha to assess internal consistency of items included in the final versions of both instruments. Following expert review, our research team consulted the field team for additional feedback. The field team suggested some minor wording changes for clarification that were incorporated into the final version (e.g. adding "while you

were in the hospital" to clarify the timeframe of reference for some questions).

Both the final Patient Delirium Burden (DEL-B-P) and Family Caregiver Delirium Burden (DEL-B-C) instruments (**Table 1**) have eight two-level questions where if the answer to the stem question (whether a certain delirium burden feature was experienced or not) was positive, a follow-up question was answered on a 0-4 scale about how upsetting/distressing the experience was (not at all upsetting/distressing - extremely upsetting/distressing) for the respondent. This format is analogous to that used in the Neuropsychiatric Inventory-Questionnaire (Cummings et al., 1994; Kaufer et al., 1998). For each pair of questions, we rated a Delirium Burden (DEL-B) item as zero if the respondent reported that the patient (or family caregiver, as relevant) did not experience the particular DEL-B item, if the respondent was uncertain or refused to answer that question. If patients or family caregivers refused to answer all questions, they were not included in analyses. For each item, a score of 1 indicates that the patient or family caregiver confirmed that they experienced the DEL-B item but it was "not at all distressing." Scores of 2-5 indicate both that they experienced the DEL-B item and that it was distressing, with higher numbers indicating greater distress. Scores were summed across the 8 paired items, yielding a minimum possible score (least burdensome) of 0 and a maximum (most burdensome) of 40. DEL-B items that were endorsed but were self-reported as "not at all distressing" contributed to the total score because awareness of these experiences is part of the subjective experience of delirium and because individuals may vary widely in their likelihood of reporting any feelings of distress.

Responses from the expert panel were generally positive for both sets of burden questions, with above-average ratings for both relevance and clarity, and clear suggestions for improvement; therefore, we did not perform a second round of review.

Testing in a target population

We assessed the DEL-B-P and DEL-B-C instruments in surgical and medical patients and their family caregivers utilizing a large subset of the Better Assessment of Illness (BASIL) study conducted at Hebrew SeniorLife and Beth Israel Deaconess Medical Center, whose Institutional Review Boards

approved this study. BASIL is an ongoing prospective, observational study of 352 surgical and medical

patients with the primary goal of developing and testing measures of delirium severity. Eligibility criteria

for patients included age ≥ 70 years, ability to communicate in English, and expected hospital stay of >2

days. We excluded patients for: imminently terminal condition; recent alcohol dependence, history of

schizophrenia or psychosis, severe deafness; and nonverbal condition (e.g., aphasic, intubated). Patients

with dementia were included. We screened participants for their capacity to consent using a standard

Capacity for Informed Consent instrument and requested written consent from all participants and/or their

proxies. Patients who were unable to communicate or interact with an interviewer were excluded from the

study, which included some patients with advanced dementia.

The BASIL study consists of: an initial in-hospital assessment; daily hospital interviews; follow-up

interviews at 5-10 days (phone), 1 and 12 months post-discharge (in-person); and chart abstraction. We

administered the final DEL-B instruments to patients and their family caregivers during the one-month

visit. To assess retest reliability, we administered the DEL-B-C to a subset of family caregivers at the in-

hospital assessment. We did not do so among patients because we did not expect reliable responses during

active delirium.

In addition to the burden instruments, we collected data on delirium status, delirium severity,

cognitive function, and other clinical characteristics and outcomes. We assessed delirium status and

severity by the Confusion Assessment Method (CAM) and CAM-Severity (CAM-S) score (long form).

The CAM (Inouye et al., 1990) is a standardized approach for detecting delirium with high sensitivity

(94-100%), specificity (90-95%) and reliability (Inouye et al., 1990; Wei, Fearing, Sternberg, & Inouye,

2008). The CAM-S is a score created from the 10 features of the full CAM instrument and ranges from 0-

19, with higher scores indicating more severe delirium(Inouye, Kosar, et al., 2014). We assessed both

CAM-S peak, the highest single CAM-S rating during hospitalization, using pre-specified cutpoints

(Vasunilashorn et al., 2018) and CAM-S sum, which is the sum of CAM-S scores on all hospital days, on

a continuous scale (Vasunilashorn et al., 2016).

To test convergent validity, we first compared DEL-B-P and DEL-B-C scores for patients who did and did not develop delirium, and second, examined the correlation between DEL-B and delirium severity. We hypothesized that DEL-B-P and DEL-B-C would be higher in patients who developed delirium and would be associated with greater delirium severity. Differences between DEL-B scores by delirium status were assessed by t-tests. We assessed associations between CAM-S sum score and DEL-B (both continuous measures) by Spearman rank correlations (rho) and CAM-S peak tertiles (categorical variable) and DEL-B by ANOVA. All analyses were conducted in StataIC (StataCorp, 2017).

Results

Cohort Characteristics

Table 4 describes the characteristics of the 267 patients from the BASIL study who completed the DEL-B-P instrument and/or the DEL-B-C instrument.

Content validity

Patients: DEL-B-P items

On average, experts deemed all patient items at least somewhat relevant (≥2 points) and at least moderately clear (≥3 points) (**Table 2**). All of the final DEL-B-P items were scored as "top 5" by at least two reviewers and had clarity and relevancy ratings ≥ 3.25 except for two items (**Table 2**). Item 4 (concerning nightmares or vivid dreams) and Item 13 (concerning physical restraints) had high clarity scores but average relevancy scores of 2.75 and 3.13, respectively. However, we retained these items because of their "top 5" ratings and clinical judgment of the study research team. Cronbach's alpha was 0.86 for the final DEL-B-P instrument, indicating good internal consistency.

Family caregivers: DEL-B-C items

On average, the experts deemed all DEL-B-C questions at least somewhat relevant (≥2 points) and at least moderately clear (≥3 points) (**Table 3**). Expert panel members expressed concern that some items would be burdensome in patients with dementia as well; thus, we prioritized the choice of final items based on specificity to delirium. All of the final DEL-B-C burden questions were scored as "top 5" by at least two reviewers and had relevancy and clarity ratings ≥3.25. Cronbach's alpha was .82 for the final DEL-B-C instrument, indicating good internal consistency.

Delirium burden in target population

Patients: DEL-B-P in target population

We assessed the DEL-B-P instrument in 247 BASIL patients who completed the 1-month interview; only one patient refused to answer all of the burden questions. The DEL-B-P instrument took three minutes on average (range 1-16 minutes) to administer. DEL-B-P scores spanned the full possible range (0-40). The median DEL-B-P score in the total sample (delirium and non-delirium combined) was 2 (interquartile range, IQR=0-7) out of 40 possible points (**Figure 1**). The median and IQR for DEL-B-P was 8 (IQR 1-14) in delirious patients compared to a median of 2 (IQR 0-5) in non-delirious patients (**Figure 2**). On average, DEL-B-P was 5.1 points (95% CI 3.3-7.0; p<.001) higher in patients who developed delirium compared to those who did not.

Family caregivers: DEL-B-C in target population

We administered the DEL-B-C instrument to 213 caregivers who completed the 1-month interview; only one caregiver refused to answer all of the burden questions. The DEL-B-C instrument took two minutes on average (range 1-10 minutes) to administer. DEL-B-C scores ranged from 0-35. The median score in the total sample was 7 (IQR=2-12) out of 40 possible points (**Figure 1**). Family caregivers of delirious patients reported a median DEL-B-C score of 12.5 (IQR 5-18) compared to a

median of 5 (IQR 1-10) in non-delirious patients (**Figure 2**). On average, DEL-B-C was 5.8 points (95%

CI 3.6-8.0; p<.001) higher in patients who developed delirium compared to those who did not.

Retest reliability of DEL-B-C

In family caregivers who completed both the in-hospital and 1-month assessments (n=143), DEL-

B-C scores were strongly correlated (Spearman rho=.73, p<.0001; intraclass correlation coefficient,

ICC=.71, 95% confidence interval, CI: .63, .78), indicating good retest reliability (Koo & Li, 2016). We

observed a trend that DEL-B-C tended to be higher at one-month post-discharge than at the in-hospital

assessment. However, a paired t-test revealed that the mean difference of 0.6 points was not significant

(95% C.I. -1.5, 0.3, p=.18).

Correlation between DEL-B-P and DEL-B-C

DEL-B-P at one month was modestly correlated with DEL-B-C at one month (Spearman's rho=.28,

p=.001). Since these correlations were lower than expected, we performed a supplemental analysis

stratified by whether or not the family caregiver lived with the patient (**Figure 3**). Correlations were

higher for patients and family caregivers who lived together (Spearman's rho=.34, p=.001) versus those

who did not (rho=.16, p=.10). However, these two correlations were not significantly different (z=1.3,

p=.20). On average, DEL-B-C was significantly higher than DEL-B-P by 2.9 points (95% CI: 1.6, 4.1);

this was true for both patients and family caregivers who lived together (difference=2.2, 95% C.I. 0.4,

4.0) and those who did not (difference=3.4, 95% C.I. 1.7, 5.1).

Associations between DEL-B and delirium severity

For both patients and family caregivers, higher delirium severity was associated with greater DEL-

B (**Figure 4**).

Patients: Associations between DEL-B-P and delirium severity

Greater DEL-B-P was correlated with greater delirium severity by CAM-S sum (Spearman's rho=.31, p<.001), indicating moderate correlation. We also observed significant differences in DEL-B-P scores between tertiles of CAM-S peak scores (F=13.1, p<.001). The average DEL-B-P score in the lowest tertile (n=103) was 2.5 (standard deviations, SD 3.2), middle tertile (n=112) was 6.2 (SD 7. 4), and highest tertile (n=32) was 7.8 (SD 9.2). Bonferroni-corrected pair-wise comparisons revealed that CAM-S peak tertiles 2 and 3 (i.e., higher CAM-S peak scores) had greater DEL-B-P scores compared to the lowest tertile (p<.005) but were not significantly different from each other (p=.64).

Family caregivers: Associations between DEL-B-C and delirium severity

Greater DEL-B-C was associated with greater patient delirium severity by CAM-S sum (Spearman's rho=.27, p<.001), indicating a small to medium correlation. We also observed significant differences in DEL-B-C scores between tertiles of CAM-S peak scores (F=9.9, p<.001). The average DEL-B-C score in the lowest tertile (n=94) was 6.2 (SD 5.6), middle tertile (n=90) was 8.1 (SD 7.1), and highest tertile (27) was 12.7 (SD 8.3). Bonferroni-corrected pair-wise comparisons revealed that CAM-S peak tertile 3 (i.e., highest CAM-S peak scores) had greater DEL-B-C compared to tertiles 1 (p<.001) and 2 (p=.007), but tertiles 1 and 2 were not different from each other (p=.12).

Discussion and Implications

We have developed and tested new delirium burden measures designed to quantify the subjective experience of delirium for patients and their family caregivers. These instruments were feasible and brief (2-3 minutes on average) to administer and acceptable to patients and their families. Following recommended guidelines (DeVellis, 2016; Lynn, 1986), we identified and operationalized content domains, evaluated and refined items with input from an interdisciplinary expert panel, and prospectively tested the instruments in an appropriate target population of medical-surgical patients and their family caregivers.

Most patients and family caregivers reported a moderate level of burden if the patient developed delirium, but little to no burden if the patient did not develop delirium. About half the patient sample overall reported DEL-B-P scores of 2 or less; and about a quarter of the family caregiver sample overall reported DEL-B-C scores of 2 or less. By contrast, half of the patient sample with delirium reported DEL-B-P scores of 8 or higher, and half of the delirium patient caregiver sample reported DEL-B-C scores of 12.5 or higher. Interestingly, patient and family caregiver DEL-B scores showed only mild correlations. However, the correlation was stronger for patients and caregivers who lived together. Overall, the results suggest that patients and their family caregivers experience the burden of delirium differently, and that a burdensome experience for a family caregiver may not translate to a burdensome experience for a patient, and vice versa. Consistent with previous reports (Partridge et al., 2013), we also found that on average, DEL-B scores were higher in family caregivers than for patients. This may be in part because the delirium itself, or contributions from cognitive impairment, may have affected patients' ability to accurately recall the event. Retest reliability of DEL-B-C was strong at one month (rho=.73); reliability of DEL-B-P could not be assessed since burden was not measured during hospitalization when a patient may have been experiencing an active delirium episode.

Finally, convergent validity was demonstrated: DEL-B scores were higher in patients who developed delirium compared to patients who did not develop delirium, and were correlated with delirium severity measures with small-moderate effect sizes. While DEL-B appears to be at least somewhat related to the severity of the delirium episode, the experience of burden is incompletely explained by delirium severity alone, suggesting that additional factors contribute to the subjective experience of delirium. For instance, burden may be driven by delirium subtype (e.g. hypoactive vs. hyperactive), emotional lability, functional impairments, incontinence, sleep disruption, comorbidity, socioeconomic or other factors which may influence the burdensome nature of the episode. An important area for future work will be to identify the predictors of delirium burden.

Strengths of this study include the robust approach to instrument development and testing, the prospective cohort utilized for testing, and the feasible measures that emerged. However, several

important caveats of our study should be noted. First, burden is a complex construct that is difficult to capture in a single instrument; thus, important items may not have been included. Since we focused on distress, aspects such as physical demands and costs of care were not assessed. Furthermore, it is possible that a different expert panel would have led to selection of different items; we cannot exclude this potential source of bias despite our efforts for cross-disciplinary representation. Although we used a rigorous approach to select the final items for our instrument, we acknowledge that these burden instruments may not be comprehensive and important contributors to the delirium experience may not have been fully captured. Second, we assessed DEL-B at one-month post-discharge, but it is quite possible that different results may have been generated at different timepoints. We found that the DEL-B-C instrument had acceptable test-retest reliability from hospitalization to one month later. The modestly higher scores at one month may have been influenced by many factors, including physical demands of caregiving not present during hospitalization, prolonged recovery period for patients, or post-traumatic stress. It is likely that even higher DEL-B scores would have been reported if patients and families had been interviewed during the peak of delirium; however, this was not feasible in the present study. The fact that so much burden is still reported at one month after the episode stresses the importance of the problem. Our data suggest that patients and family caregivers would benefit from support not only during and immediately after a delirium episode, but also some time after when they have had time to reflect on their experience. Future studies of test-retest at different intervals in both patients and caregivers are needed to better understand the reliability of these instruments. Third, the self-report of distress may differ by study population and factors such as age, sex, race, ethnicity, and country of origin (Knight & Sayegh, 2010; Napoles, Chadiha, Eversley, & Moreno-John, 2010; Pinquart & Sorensen, 2005; Scharlach et al., 2006). Moreover, burden may differ between patients with hyperactive compared to hypoactive delirium, with underlying dementia, or across different settings (e.g. surgery, intensive care, palliative care). Finally, it is likely that patients who cannot recall their delirium—due to delirium itself or other causes of cognitive impairment like dementia—will inaccurately report delirium burden. This is an important area for future investigation. Because our baseline assessment occurred after hospital

admission, delirium and other factors could influence cognitive functioning, making it difficult to determine a patient's true cognitive baseline. Future studies should further investigate whether pre-hospitalization factors, like dementia severity, are associated with delirium burden. In addition, future studies will be critical to examine the generalizability of these measures to other populations and settings.

Delirium is often distressing to patients and caregivers. Development of tools to measure different features contributing to delirium burden is a critical step towards improving support and management of delirium for patients and their family caregivers. Additional validation studies of these instruments are needed to examine predictive validity for important outcomes, such as quality of life, physical and emotional health, mortality, healthcare costs for both patients and their family caregivers, and to assure generalizability of these measures across diverse study populations and settings. While interventions to reduce the risk of and severity of delirium should reduce delirium burden, our results suggest that interventions may need to focus more specifically on improving the lived experience of delirium for patients and their family caregivers. Burden from delirium is a tremendously important element to consider in the clinical management of delirium. Sustained emotional and psychological consequences from delirium are being increasingly recognized (Drews et al., 2015; Langan et al., 2017). Thus, burden should be carefully addressed as another important dimension of delirium—and ultimately, considered as an outcome for clinical trials and other intervention studies for delirium. We hope these instruments will lay the foundation for this important future work.

References

Breitbart, W., Gibson, C., & Tremblay, A. (2002). The delirium experience: Delirium recall and delirium-related distress in hospitalized patients with cancer, their spouses/caregivers, and their nurses. *Psychosomatics, 43*(3), 183-194. doi: 10.1176/appi.psy.43.3.183

Bruera, E., Bush, S.H., Willey, J., Paraskevopoulos, T., Li, Z., Palmer, J.L., . . . Elsayem, A. (2009). Impact of delirium and recall on the level of distress in patients with advanced cancer and their family caregivers. *Cancer, 115*(9), 2004-2012. doi: 10.1002/cncr.24215

Chiao, C.Y., Wu, H.S., & Hsiao, C.Y. (2015). Caregiver burden for informal caregivers of patients with dementia: A systematic review. *Int Nurs Rev, 62*(3), 340-350. doi: 10.1111/inr.12194

Cummings, J.L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D.A., & Gornbein, J. (1994). The neuropsychiatric inventory: Comprehensive assessment of psychopathology in dementia. *Neurology, 44*(12), 2308-2314.

DeVellis, R.F. (2016). *Scale development: Theory and applications* (Vol. 26): Sage publications.

Drews, T., Franck, M., Radtke, F.M., Weiss, B., Krampe, H., Brockhaus, W.R., . . . Spies, C.D. (2015). Postoperative delirium is an independent risk factor for posttraumatic stress disorder in the elderly patient: A prospective observational study. *Eur J Anaesthesiol, 32*(3), 147-151. doi: 10.1097/EJA.0000000000000107

Finucane, A.M., Lugton, J., Kennedy, C., & Spiller, J.A. (2017). The experiences of caregivers of patients with delirium, and their role in its management in palliative care settings: An integrative literature review. *Psychooncology, 26*(3), 291-300. doi: 10.1002/pon.4140

Fuller, V. (2016). Delirium recall - an integrative review. *J Clin Nurs, 25*(11-12), 1515-1527. doi: 10.1111/jocn.13155

Grover, S., Ghosh, A., & Ghormode, D. (2015). Experience in delirium: Is it distressing? *J Neuropsychiatry Clin Neurosci, 27*(2), 139-146. doi: 10.1176/appi.neuropsych.13110329

Inouye, S.K., Kosar, C.M., Tommet, D., Schmitt, E.M., Puelle, M.R., Saczynski, J.S., . . . Jones, R.N. (2014). The cam-s: Development and validation of a new scoring system for delirium severity in 2 cohorts. *Ann Intern Med, 160*(8), 526-533. doi: 10.7326/M13-1927

Inouye, S.K., van Dyck, C.H., Alessi, C.A., Balkin, S., Siegal, A.P., & Horwitz, R.I. (1990). Clarifying confusion: The confusion assessment method. A new method for detection of delirium. *Ann Intern Med, 113*(12), 941-948.

Inouye, S.K., Westendorp, R.G., & Saczynski, J.S. (2014). Delirium in elderly people. *Lancet, 383*(9920), 911-922. doi: 10.1016/S0140-6736(13)60688-1

Kaufer, D.I., Cummings, J.L., Christine, D., Bray, T., Castellon, S., Masterman, D., . . . DeKosky, S.T. (1998). Assessing the impact of neuropsychiatric symptoms in alzheimer's disease: The neuropsychiatric inventory caregiver distress scale. *J Am Geriatr Soc, 46*(2), 210-215.

Knight, B.G., & Sayegh, P. (2010). Cultural values and caregiving: The updated sociocultural stress and coping model. *J Gerontol B Psychol Sci Soc Sci, 65B*(1), 5-13. doi: 10.1093/geronb/gbp096

Koo, T.K., & Li, M.Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med, 15*(2), 155-163. doi: 10.1016/j.jcm.2016.02.012

Langan, C., Sarode, D.P., Russ, T.C., Shenkin, S.D., Carson, A., & Maclullich, A.M.J. (2017). Psychiatric symptomatology after delirium: A systematic review. *Psychogeriatrics, 17*(5), 327-335. doi: 10.1111/psyg.12240

Lutomski, J.E., Baars, M.A., Schalk, B.W., Boter, H., Buurman, B.M., den Elzen, W.P., . . . Consortium, T.-M. (2013). The development of the older persons and informal caregivers survey minimum dataset (topics-mds): A large-scale data sharing initiative. *PLoS One, 8*(12), e81673. doi: 10.1371/journal.pone.0081673

Lynn, M.R. (1986). Determination and quantification of content validity. *Nursing research, 35*(6), 382-386.

Marcantonio, E.R. (2018). Delirium in hospitalized older adults. *N Engl J Med, 378*(1), 96-97. doi: 10.1056/NEJMc1714932

Morandi, A., Rogers, B.P., Gunther, M.L., Merkle, K., Pandharipande, P., Girard, T.D., . . . Visions Investigation, V.I.S.N.S. (2012). The relationship between delirium duration, white matter integrity, and cognitive impairment in intensive care unit survivors as determined by diffusion tensor imaging: The visions prospective cohort magnetic resonance imaging study*. *Crit Care Med, 40*(7), 2182-2189. doi: 10.1097/CCM.0b013e318250acdc

Napoles, A.M., Chadiha, L., Eversley, R., & Moreno-John, G. (2010). Reviews: Developing culturally sensitive dementia caregiver interventions: Are we there yet? *Am J Alzheimers Dis Other Demen, 25*(5), 389-406. doi: 10.1177/1533317510370957

Novak, M., & Guest, C. (1989). Application of a multidimensional caregiver burden inventory. *Gerontologist, 29*(6), 798-803.

Partridge, J.S., Martin, F.C., Harari, D., & Dhesi, J.K. (2013). The delirium experience: What is the effect on patients, relatives and staff and what can be done to modify this? *Int J Geriatr Psychiatry, 28*(8), 804-812. doi: 10.1002/gps.3900

Pinquart, M., & Sorensen, S. (2005). Ethnic differences in stressors, resources, and psychological outcomes of family caregiving: A meta-analysis. *Gerontologist, 45*(1), 90-106.

Scharlach, A.E., Kellam, R., Ong, N., Baskin, A., Goldstein, C., & Fox, P.J. (2006). Cultural attitudes and caregiver service use: Lessons from focus groups with racially and ethnically diverse family caregivers. *J Gerontol Soc Work, 47*(1-2), 133-156. doi: 10.1300/J083v47n01_09

Schmitt, E.M., Gallagher, J., Albuquerque, A., Tabloski, P., Lee, H.J., Gleason, L., . . . Schulman-Green, D. (2017). Perspectives on the delirium experience and its burden: Common themes among older patients, their family caregivers, and nurses. *The Gerontologist*, gnx153-gnx153. doi: 10.1093/geront/gnx153

StataCorp. (2017). Stata statistical software: Release 15 (Version 15). College Station, TX: StataCorp LLC.

Vasunilashorn, S.M., Fong, T.G., Albuquerque, A., Marcantonio, E.R., Schmitt, E.M., Tommet, D., . . . Inouye, S.K. (2018). Delirium severity post-surgery and its relationship with long-term cognitive decline in a cohort of patients without dementia. *J Alzheimers Dis, 61*(1), 347-358. doi: 10.3233/JAD-170288

Vasunilashorn, S.M., Marcantonio, E.R., Gou, Y., Pisani, M.A., Travison, T.G., Schmitt, E.M., . . . Inouye, S.K. (2016). Quantifying the severity of a delirium episode throughout hospitalization: The combined importance of intensity and duration. *J Gen Intern Med, 31*(10), 1164-1171. doi: 10.1007/s11606-016-3671-9

Wei, L.A., Fearing, M.A., Sternberg, E.J., & Inouye, S.K. (2008). The confusion assessment method: A systematic review of current usage. *J Am Geriatr Soc, 56*(5), 823-830. doi: 10.1111/j.1532-5415.2008.01674.x

Zarit, S.H., Reever, K.E., & Bach-Peterson, J. (1980). Relatives of the impaired elderly: Correlates of feelings of burden. *Gerontologist, 20*(6), 649-655.

FIGURE LEGENDS

Figure 1. Histograms of patient delirium burden (DEL-B-P, left) and family caregiver delirium burden (DEL-B-C, right) instruments assessed at one month post-hospitalization. The total sample is displayed in the top panel, and the subsample of delirium patients only are displayed on the bottom panel. All data displayed is from the one-month assessment.

Figure 2. Delirium burden (DEL-B) by delirium status for patients (left) and family caregivers (right). Box plots display Patient Delirium Burden (DEL-B-P) and Family Caregiver Delirium Burden (DEL-B-C) data by delirium status (no delirium vs. delirium) for data collected at the 1-month assessment.

Figure 3. Correlation between Family Caregiver Delirium Burden (DEL-B-C) and Patient Delirium Burden (DEL-B-P). The correlation between DEL-B-P and DEL-B-C was stronger for caregiver-patient dyads who lived together (blue triangles/solid line) than for caregiver-patient dyads who did not live together (red circles/dashed line). Rho refers to Spearman's rank order correlation coefficient.

Figure 4. Association between delirium severity and patient (top) and family caregiver (bottom) delirium burden (DEL-B). Delirium severity was measured by CAM-S summed score (left, scatter plot) and CAM-S peak tertiles (right, box plot).

TABLES AND FIGURES

Table 1. Patient Delirium Burden (DEL-B-P) and Family Caregiver Burden Delirium (DEL-B-C)
Instrument Elements[a]

Patient items (DEL-B-P)
•   Unsure of where they were
•   Could not remember parts of the hospital stay
•   Saw or heard things that were not really there
•   Nightmares or vivid dreams that were intense or bothersome
•   Suddenly felt confused
•   Thought that they would not get better
•   Afraid of losing their mind
•   Restricted from getting out of a bed or a chair with alarms or restraints
Family caregiver items (DEL-B-C)
•   Loved one did not recognize caregiver
•   Loved one experienced changes in memory and thinking
•   Loved one saw or heard things that were not really there
•   Loved one became irritable or angry
•   Feelings of helplessness as a caregiver
•   Concern about increased responsibilities as a caregiver
•   Concern that loved one would never be back to his/her usual self
•   Loved one demonstrated unsafe behaviors

[a]The instruments are copyright pending and will be made freely and publicly available at
https://www.hospitalelderlifeprogram.org/delirium-instruments/.

Table 2. Expert Ratings of Patient Delirium Burden (DEL-B-P) Items[a]

Burden theme	Items for expert review	Relevance	Clarity	Top 5[b]	Included/Reason not Included
Symptom Burden	1. Unsure of where they were	3.8	3.6	4 (50%)	Included
	2. Could not remember parts of the hospital stay	3.4	3.9	2 (25%)	Included
	3. Saw or heard things that were not really there	3.6	3.9	3 (38%)	Included
	4. Nightmares or vivid dreams that were intense or bothersome	2.8	3.6	2 (25%)	Included
	5. Difficulty communication	2.9	3.0	1 (13%)	Dropped due to low relevancy rating
	6. Suddenly felt confused	3.9	3.9	6 (75%)	Included
Emotional Burden	7. Felt they may never be normal again	2.9	3.1	1 (13%)	Dropped due to low relevancy rating
	8. Thought that they would not get better	3.3	3.8	2 (25%)	Included
	9. Afraid of losing their mind	3.8	3.6	6 (75%)	Included
Situational Burden	10. Felt the hospital staff did not listen to them	2.6	3.0	1 (13%)	Dropped due to low relevancy rating
	11. Felt they were being scolded by hospital staff	2.4	3.4	0	Dropped due to low relevancy rating
	12. Felt they were not being respected by hospital staff	2.9	3.4	2 (25%)	Dropped due to low relevancy rating
	13. Restricted from getting out of a bed or a chair with alarms or restraints	3.1	3.5	4 (50%)	Included
	14. Concerned or being/becoming a burden to others	3.1	3.8	2 (25%)	Dropped due to non-specificity to delirium
	15. Concerned about how they would function in the future	3.0	3.4	2 (25%)	Dropped due to non-specificity to delirium

[a]The instruments are copyright pending and will be made freely and publicly available at https://www.hospitalelderlifeprogram.org/delirium-instruments/.
[b]Top 5 refers to the n (%) of reviewers who selected the item in response to the question "If you were to develop a 5 item delirium burden questionnaire based on the above items, which would you choose?"

Table 3. Expert Ratings of Family Caregiver Delirium Burden (DEL-B-C) Items[a]

Burden Theme	Items for expert review	Relevance	Clarity	Top 5[b]	Included/Reason not Included
Symptom Burden	1. Loved one did not recognize caregiver	3.5	4.0	2 (25%)	Included
	2. Loved one behaved in an upsetting way	2.5	3.1	0	Dropped due to low relevancy rating
	3. Loved one saw or heard things that were not really there	3.6	4.0	2 (25%)	Included
	4. Loved one did not know where he/she was	3.5	3.9	1 (13%)	Dropped due to length
	5. Loved one experienced changes in memory and thinking (wording 1)	3.8	3.8	2 (25%)	Included
	6. Loved one experienced changes in memory and thinking (wording 2)	3.8	3.4	2 (25%)	Dropped due to comparable item (#5)
	7. Loved one became irritable or angry	3.0	3.5	3 (38%)	Included
	8. Loved one had troubles understanding	2.8	3.5	1 (13%)	Dropped due to low relevancy rating
	9. Difficulty understanding loved one	3.0	3.4	0	Dropped due to non-specificity to delirium
	10. Difficulty communicating with loved one	2.6	3.4	1 (13%)	Dropped due to low relevancy rating
	11. Feelings that his/her presence made loved one's symptoms worse	2.9	3.4	1 (13%)	Dropped due to low relevancy rating
Emotional Burden	12. Feelings of helplessness as a caregiver (wording 1)	3.3	3.4	3 (38%)	Dropped due to comparable item (#13)
	13. Feelings of helplessness as a caregiver (wording 2)	3.6	3.5	6 (75%)	Included
	14. Concern about loved one's future	3.1	3.0	0	Dropped due to non-specificity to delirium
	15. Concern about increased responsibilities as a caregiver	3.9	3.8	4 (50%)	Included
	16. Concern that loved one would lose their independence	3.1	3.0	0	Dropped due to non-specificity to delirium
	17. Loved one refused his/her help	2.0	3.6	0	Dropped due to low relevancy rating
	18. Thoughts that loved one would not recover	3.1	3.3	0	Dropped due to non-specificity to delirium
	19. Concern that loved one would never be back to his/her usual self	3.7	3.3	2 (25%)	Included
Situational Burden	20. Concern that the doctors couldn't fix what was wrong	2.9	3.4	1 (13%)	Dropped due to low relevancy rating
	21. Loved one demonstrated unsafe behaviors	3.8	3.8	5 (63%)	Included

	22. Feelings that loved one was not being taken seriously by the hospital staff	2.8	3.1	1 (13%)	Dropped due to low relevancy rating

[a]The instruments are copyright pending and will be made freely and publicly available at https://www.hospitalelderlifeprogram.org/delirium-instruments/.
[b]Top 5 refers to the n (%) of reviewers who selected the item in response to the question "If you were to develop a 5 item delirium burden questionnaire based on the above items, which would you choose?"

Table 4. Burden Sample Patient Characteristics

Patient Characteristics	Delirium burden Sample (N=267)
Age – mean (SD)	80.3 (6.8)
Sex – n (% female, self-reported)	152 (57%)
Nonwhite – n (%)	45 (17%)
Education – mean (SD)	14.6 (2.9)
Charlson score – n (%)	
0	71 (27%)
1	57 (21%)
2+	139 (52%)
Dementia – n (%)	79 (30%)
Any ADL impairment – n (%)	204 (77%)
Any IADL impairment – n (%)	108 (60%)
Delirium – n (%)	59 (22%)
CAM-S peak score – n (%)	
0-2	113 (42%)
3-7	121 (45%)
8-19	33 (12%)
CAM-S sum score – mean (SD)	9.0 (9.4)
Caregiver relationship to patient – n (%)	
Spouse/Partner	75 (32%)
Son/Daughter	112 (47%)
Son/Daughter in law	3 (1%)
Grandchild	6 (3%)
Brother/Sister	14 (6%)
Nephew/Niece	9 (4%)
Cousin	1 (<1%)
Other relative	1 (<1%)
Friend/Neighbor	16 (7%)
Paid employee or caretaker	1 (<1%)
Caregiver lives with patient – n (%)	109 (45%)

Patient delirium burden and Caregiver delirium burden samples are partially overlapping; 247 patients and 213 family caregivers completed the burden assessment at Month 1; 193 patient-caregiver dyads completed both burden instruments. Missing data: Information on education was only available in 264 patients, any ADL impairment in 264, any IADL impairment in 181, Caregiver relationship to patient in 238, and Caregiver lives with patient in 238. All other variables had no missing data for the 267 patients with patient and/or caregiver burden data. ADL = Activities of Daily Living, IADL = Instrumental Activities of Daily Living, Dementia = proxy Informant Questionnaire of Decline in Elderly (IQCODE) score ≥ 3.5 or chart-based diagnoses of Alzheimer's disease or dementia.
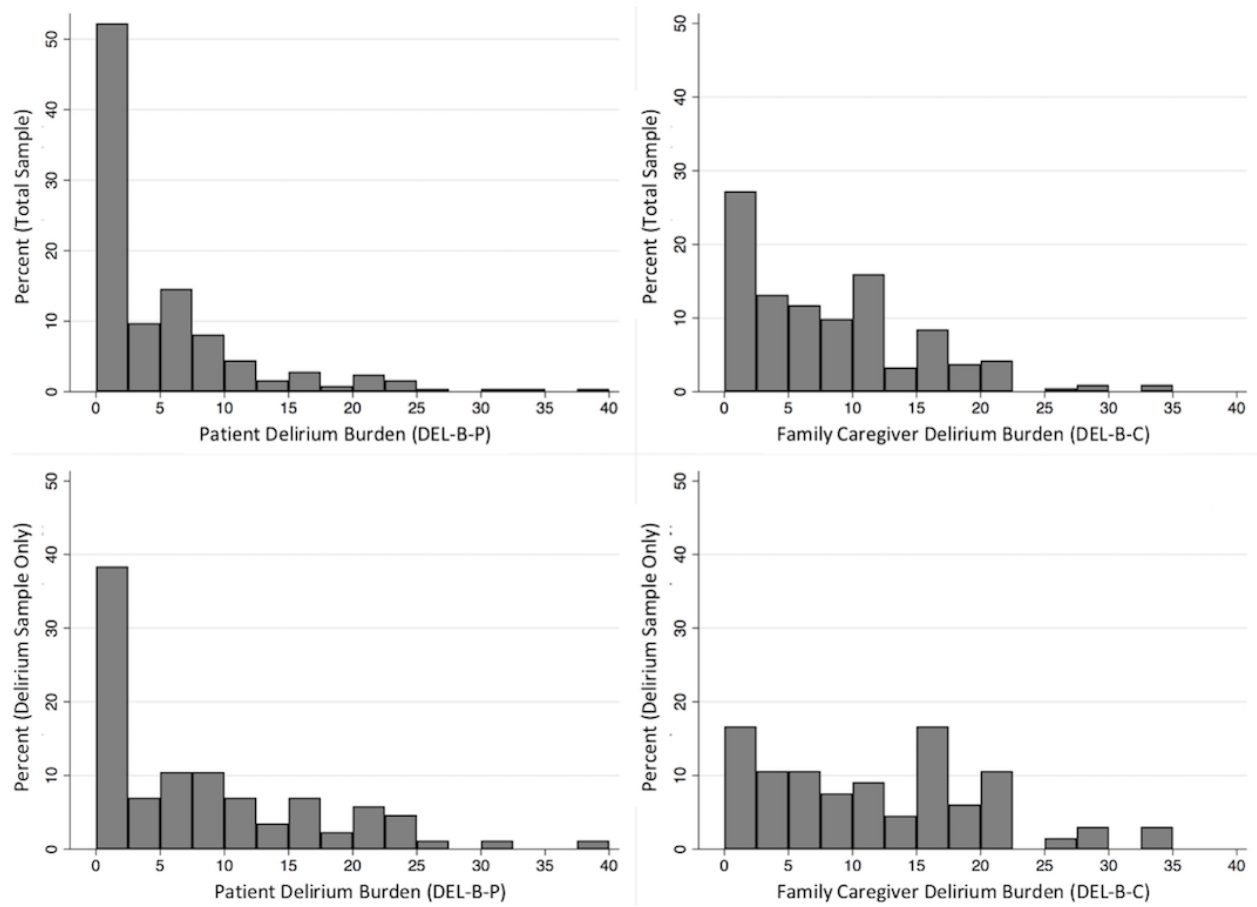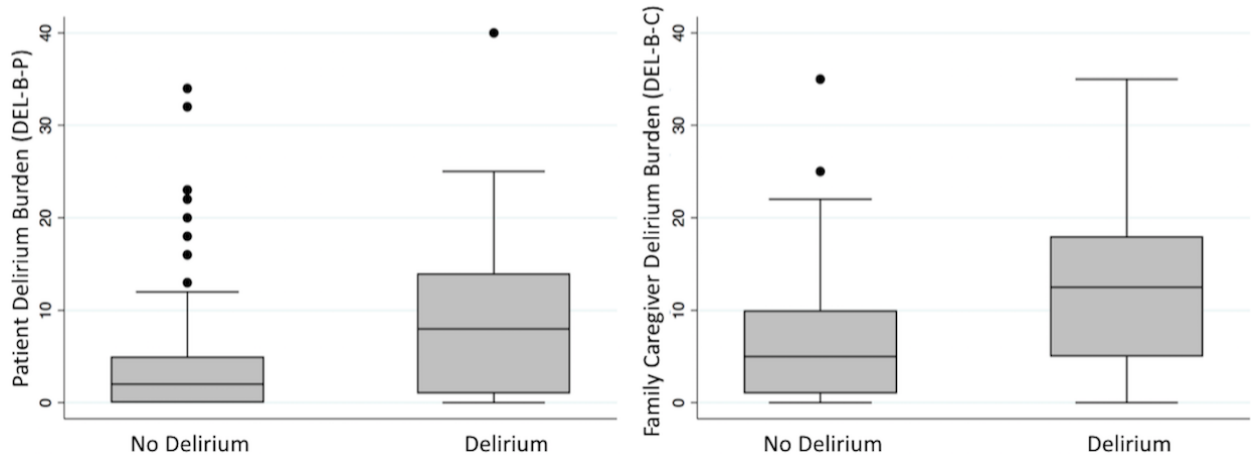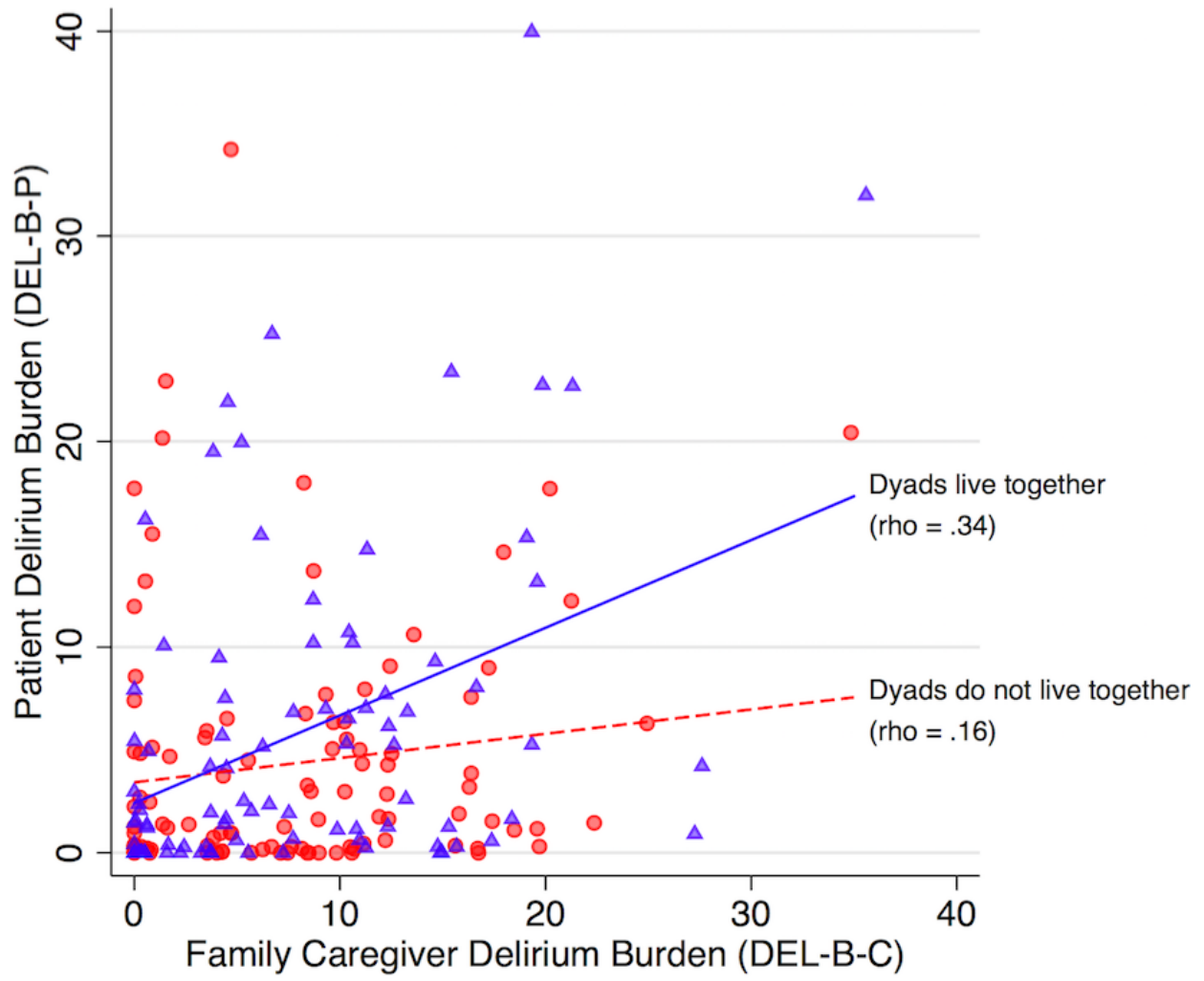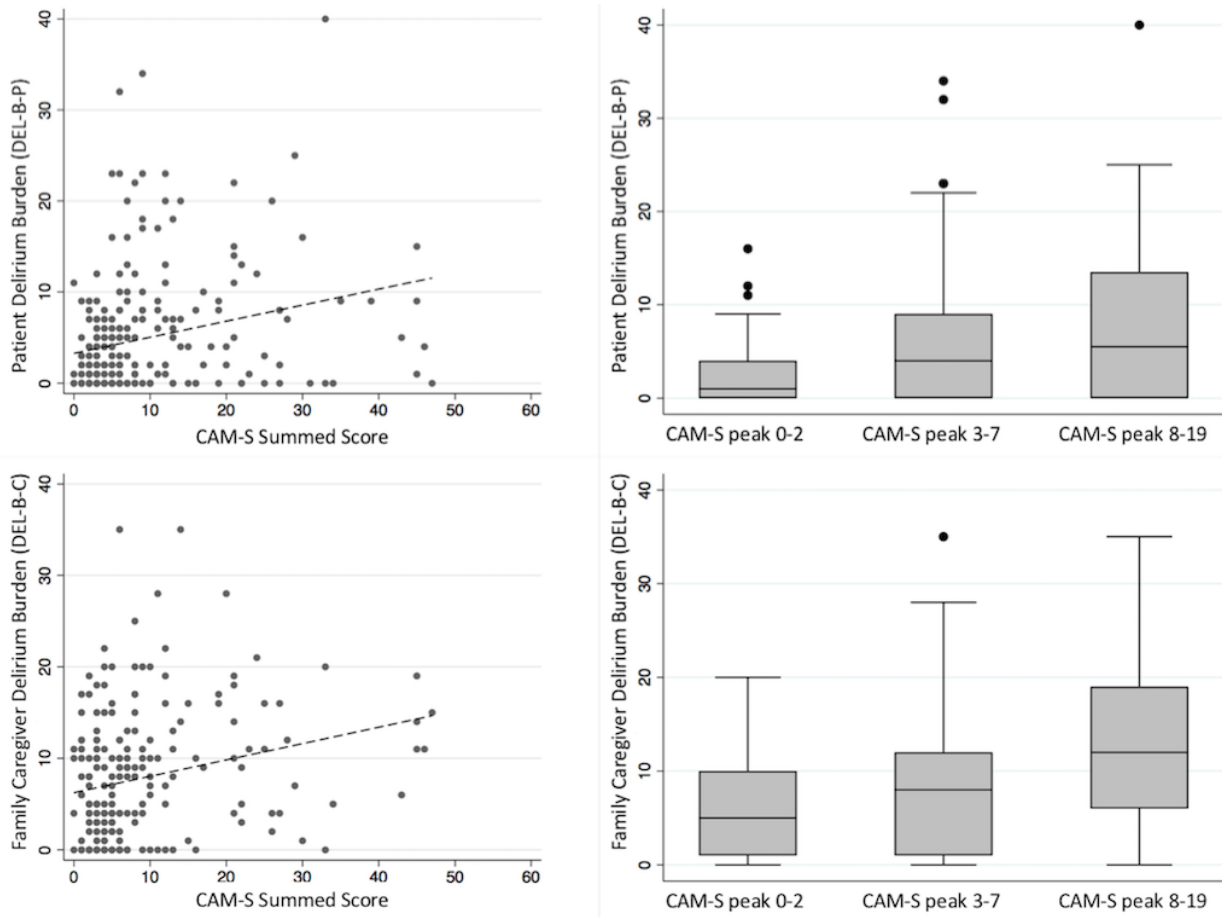
Figure 1.

Figure 2.

Figure 3.

Figure 4.

SUPPLMENTARY MATERIAL

Supplementary Table 1. Description of expert panel

Panel Characteristics	Expert panel (N=8)
Disciplines/expertise (may overlap)	Anesthesiology, critical care, geriatrics, hospital medicine, internal medicine, neurology, nursing, occupational therapy, psychiatry, surgery
Degree[a] – n (%)	
MD	4 (50%)
PhD or PsyD	2 (25%)
BS	2 (25%)
RN	1 (13%)
MPH	2 (25%)
Gender – n (%) female	5 (63%)
Geographic Locations – n (%)	
Northeast	5 (63%)
Southeast	1 (13%)
Southwest	2 (25%)

[a]Numbers do not sum to 8 because some panelists had multiple degrees and specialties

Supplementary Figure 1. Content Validity Rating Form for Patient Delirium Burden

Experiences of Delirium: Assessing Patient Burden
Content Validity Rating Form

Instructions for Expert Reviewers.   The following items are under consideration for inclusion in a delirium burden instrument to measure the burden of different delirium related experiences of patients with delirium. Each question will be answered yes or no, and a follow-up question will be asked if the caregiver endorses a yes about how upsetting the symptom was. The follow-up question will be rated on a 5-point Likert scale: 0 – Not at all upsetting, 1- A little bit upsetting, 2- Somewhat upsetting, 3- Very upsetting, 4- Extremely upsetting.

The questions will be asked for the time period of the index hospitalization until 2 weeks after discharge. Some of these questions may seem repetitious; please rate each independently. You can state your preference in the suggestions section.

For each statement, please follow these steps:

1) Evaluate the RELEVANCE of each item to measure the delirium experience of patients by circling one of the following responses:
   - 1 = not relevant (NR)
   - 2 = somewhat relevant (SR)
   - 3 = quite relevant (QR)
   - 4 = very relevant (VR)

2) Evaluate the CLARITY of each item by circling one of the following responses:
   - 1=not clear (NC)
   - 2=somewhat clear (SC)
   - 3=quite clear (QC)
   - 4=very clear (VC)

3) Make any suggestions on how we can improve the statement using the SUGGESTIONS column.

4) Make any suggestions for additional statements in the space provided on the last page of the content validity rating form.

THANK YOU FOR SHARING YOUR EXPERTISE

Supplementary Figure 2. Content Validity Rating Form for Family Caregiver Delirium Burden

Experiences of Delirium: Assessing Caregiver Burden
Content Validity Rating Form

Instructions for Expert Reviewers.   We have completed a series of qualitative interviews with caregivers, and have identified some key themes highlighted by caregivers. Based on these themes, the following items are under consideration for inclusion in a delirium burden instrument to measure the burden of different delirium related experiences on caregivers of patients with delirium. Each question will be answered yes or no, and a follow-up question will be asked if the caregiver endorses a yes about how upsetting the symptom was. The follow-up question will be rated on a 5 point Likert scale; 0 – Not at all upsetting, 1- A little bit upsetting, 2- Somewhat upsetting, 3- Very upsetting, 4- Extremely upsetting.

The questions will be asked for the time period of the index hospitalization until 2 weeks after discharge. Some of these questions may seem repetitious; please rate each independently. You can state your preference in the suggestions section. This will take about 10-15 minutes. We would appreciate your feedback by February 5, 2016.

For each item, please follow these steps:
1) Evaluate the RELEVANCE of each item to the delirium experience of caregivers by clicking the checkbox for one of the following responses:
    1 = not relevant (NR)
    2 = somewhat relevant (SR)
    3 = quite relevant (QR)
    4 = very relevant (VR)

2) Evaluate the CLARITY of each item by clicking the checkbox for one of the following responses:
    1=not clear (NC)
    2=somewhat clear (SC)
    3=quite clear (QC)
    4=very clear (VC)

3) Make any suggestions on how we can improve the item using the SUGGESTIONS column.

4) Make any suggestions for additional items in the space provided on the last page of the content validity rating form.

THANK YOU FOR SHARING YOUR EXPERTISE