**Cognition**

**Categorical/dichotomous**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Studies | Data type | Cognitive change  (decline vs no-decline) | Cognitive performance  (impairment vs no-impairment) | Note |
| Bickel 2008 | Cat |  | x | ?? Complaints about memory decline after surgery, but no information on whether it fell below the threshold or represented a change from baseline*. – Agree to put in Performance* |
| Bryson 2011 | Cat | x |  | Is this a categorical/ dichotomous change data?? I took it to change “decline vs non- decline” section  Cognitive dysfunction at three months. “A patient was classified as having POCD if the Z scores on either two individual tests or the combined Z score was -1.96 or more negative” *– clearly a change measure from baseline, the change measured as at least a 1 SD decline or more.* |
| Bulic 2020 | Cat |  | x | Ok to me. “Lower scores indicate poorer performance (< 20 = severe impairment; 20–25 = mild impairment; 26–32 = ambiguous; > 32 = normal)” *– ok, looks like cross-sectional* |
| Bulic 2020 | Cat |  | x |
| Daiello 2019 | Cat | x |  | Cognitive dysfunction. I took it to categorical decline analysis. Is this a decline? This the same as Bryson 2011 shall we take it that to decline as well. “Postoperative cognitive dysfunction was defined based on change from baseline; a composite z score of at least 1.96 across all tests, or z scores for two or more tests scores at least 1.96.” - S*ame as Bryson, a threshold used to classify as decline from baseline score.* |
| Daiello 2019 | Cat | x |  |
| Daiello 2019 | Cat | x |  |
| DelaVarga-Martínez 2022 | Cat |  | x | Is this a decline data? Memory problem, concentration problem, confusion/ disorientation. “we examined the change in cognitive function from before surgery  to 3 years after hospital discharge.” Except this statement there is no clear information on how the data categorized. *– It looks like they have two cross-sectional analyses, one at baseline and one at follow up, and are comparing the numbers, but they have not reported the actual decline as ‘baseline-follow-up’ scores. They are not categorising participants based on whether they decline or not, so not a decline.* |
| DelaVarga-Martínez 2022 | Cat |  | x |
| DelaVarga-Martínez 2022 \* Possibly ask Hannah | Cat |  | x |
| Dostovic 2021 | Cat |  | x | Ok to me. “Cognition preservation has a score of  0, and the maximum disorder has a score of 37.” |
| Franck 2016 | Cat |  | x | ??Cognitive dysfunction. I couldn’t get enough information on how they classified cognitive dysfunction. *This one is quite unclear, they have not described how the cognitive results were scored or classified any thresholds or described formula for calculating change. It states cognitive testing was done at baseline, and 1 week/3 months, but without further information, we’ll default to put it in Performance.* |
| Franck 2016 | Cat |  | x |
| Franck 2016 | Cat |  | x |
| Furlaneto 2007 | Cat | x |  | Cognitive loss. I took that to categorical change analysis section. “functional and cognitive status, we used repeated measures analysis of variance to compare the baseline and follow-up ADL, IADL, and BDRS of individual patients…..”allowing us to compare the  changes in functional and cognitive performances between cases and controls. *Reasonably clear that they categorised participants based on their ‘loss’ from baseline to follow-up.* |
| Gonçalves 2023 | Cat |  | x | Ok to me. “a 10-CS harmonized score  <=5 to define cognitive impairment. – *Clearly a cross sectional analysis at follow-up.* |
| Hempenius 2016 | Cat | x |  | Took to categorical change analysis section. Decreased cognitive function – *definitely decline, very clear explanation of what categorical decline is.* |
| Inouye 2016 | Cat | x |  | I’m thinking cognitive decline. ??? moved to cognitive change categorical section . *Based on the description in the results “if we consider only those patients who had low IQCODE ratings at baseline, 35% of those who developed delirium were above the 3.2 thresholds at 36 months vs 19% of the non-delirium group” – this feels like a cross-sectional analysis at 36 months, rather than a categorisation based on change. But, check the statistics description, in methods, it mentions a linear regression for changes in IQCODE, with time\*delirium interaction* |
| Kat 2008 | Cat |  | x | ??No clear information except mentioning they used “Peterson criteria” for assessing MCI. - *No evidence of change score, just reported the follow-up numbers based on the follow-up assessments.* |
| Koster 2012 | Cat |  | x | It is ok but self-report “We determined problems with memory, concentration, and confusion were present when the patient felt, thought, or believed he or she had a problem in this area. Memory problem, Concentration problem, Confusion - *clear that this is performance, assessment done at follow-up and not compared to a baseline ‘score’ or data.* |
| Koster 2012 | Cat |  | x |
| Koster 2012 | Cat |  | x |
| Mathies 2020 | Cat |  | x | Not that clear but I’m in favour of impairment. Neurodegeneration. *– Not much information given, it states both test performance at follow-up AND changes compared to baseline were used to categorise, BUT no further info provided. Err on the side of caution, and assumed final data was based on follow-up scores (cross-sectional).* |
| McCusker 2014 | Cat | x |  | I took that to categorical change analysis section. – *Quite clear description of the categorical cut-offs used based on change in score from baseline to follow-up, so decline* |
| Nguyen 2018 | Cat |  | x | It is okay. “An individual was defined as being cognitively impaired on the basis of the MIS criteria if they scored 4 or less. Cognitive impairment based on the CFT screen  was defined as a score of 12 or less.” – *Thresholds provided for follow-up scores rather than change thresholds.* |
| Nguyen 2018 | Cat |  | x |
| Rudolph 2008 | Cat | x |  | It is okay. I just move this to change categorical section “The ISPOCD study  identified patients with POCD using a definition based on change from baseline which consisted of either a composite Z-score of > 2 across tests or two or more tests with Z-scores > 2. “Postoperative cognitive dysfunction was defined as a composite Z-score > 2 across tests or at least two individual test Z-scores > 2.” *Similar to previous studies above, this is clearly explained way to calculate change from baseline using SD of Z score.* |
| Saczynski 2012 | Cat | x |  | Score below the baseline. Moved to change categorical section *– Table 3, data is percent of participants whose scores were below their baseline scores for MMSE, so this is a decline measure* |
| Saczynski 2012 | Cat | x |  |
| Saczynski 2012 | Cat | x |  |
| VanderHeijden 2023 | Cat |  | x | It is ok. “Scores were transformed to a 0–100 range and a total of >43 points was considered  as cognitive impairment” *– Confident this is x-sectional* |
| VanderHeijden 2023 | Cat |  | x |
| Vasunilashorn 2018 | Cat |  | x | It is ok. “IQCODE≥3.2 was used to indicate  Impairment” *– as above, data presented as n(%) of participants categorised based on their scores at each time point during follow-up.* |
| Vasunilashorn 2018 | Cat |  | x |
| Vasunilashorn 2018 | Cat |  | x |
| Vasunilashorn 2018 | Cat |  | x |
| Vasunilashorn 2018 | Cat |  | x |
| Vasunilashorn 2018 | Cat |  | x |
| Verloo 2016 | Cat |  | x | It is ok. “The sum of the scores varies from 0 (severe cognitive impairment) to 30 (no cognitive impairment). A score of, <24 points was considered as the cutoff point for cognitive impairment.” *(This is the 2-day follow-up study), and very clear Performance data.* |
| Vives-Borrás 2019 | Cat | x |  | Check! The categorical data was decline whereas the continuous data was performance. Cognitive decline. Took to cognitive decline section. *– Table 4, shows clearly that each participant was categorised based on the change in their scores from baseline, and the n(%) of these categories is presented, so decline.* |
| Wolters 2014 | Cat |  | x | ?? it reports mild/severe problems with cognitive functioning “For assessment of problems with cognitive functioning we used the sixth question of the EQ-6D questionnaire. Problems with cognitive functioning were subdivided into no problems, mild problems and severe problems. *– Can see no mention of change from baseline, it only talks about the classification based on their follow-up results.* |
| Wolters 2014 | Cat |  | x |
| Van Rijsbergen 2011 | Cat |  | x | It is ok. “lower z-score indicates a poorer performance. domain-specific disorder was considered to be present when the z-score was lower than −1.65.” ­– *This seems to be very clear, it is a comparison of z scores at follow-up between delirium and no delirium groups.* |
| Van Rijsbergen 2011 | Cat |  | x |
| Van Rijsbergen 2011 | Cat |  | x |
| Van Rijsbergen 2011 | Cat |  | x |
| Van Rijsbergen 2011 | Cat |  | x |
| Van Rijsbergen 2011 | Cat |  | x |
| Van Rijsbergen 2011 | Cat |  | x |
| Chen 2017 | Cat |  | x | It is ok. “Patients were defined as having cognitive impairment if the total score of TICS‐m  was less than 33 points.”  *Very clear, talks about results in terms of incidence at each follow up timepoint.* |

**Cognition**

**Continuous**

| Studies | Data type | Score at follow-up  (Similar to categorical performance) | Change from the baseline  (Similar to categorical decline) | Note |
| --- | --- | --- | --- | --- |
| Wolters 2017 | cont | x |  | Cognitive performance score |
| Wolters 2017 | cont | x |  |
| Witlox 2013 | cont | x |  | Cognitive performance score |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Witlox 2013 | cont | x |  |
| Whittamore 2014 | cont | x |  | Cognitive performance score |
| Vives-Borrás 2019 | cont | x |  | Cognitive performance score |
| Verloo 2016 | cont | x |  | Cognitive performance score |
| Vasunilashorn 2018 | cont | x |  | Cognitive performance score |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| Vasunilashorn 2018 | cont | x |  |
| VandenBoogaard 2012 | cont | x |  | Cognitive performance score |
| Svenningsen 2014 | cont | x |  | Cognitive performance score |
| Svenningsen 2014 | cont | x |  |
| Svenningsen 2014 | cont | x |  |
| Svenningsen 2014 | cont | x |  |
| Svenningsen 2014 | cont | x |  |
| Svenningsen 2014 | cont | x |  |
| Sheng 2006 | cont | x |  | Cognitive performance score |
| Sheng 2006 | cont | x |  |
| Serrano-Duenas 2005 | cont | x |  | Cognitive performance score |
| Serrano-Duenas 2005 | cont | x |  |
| Serrano-Duenas 2005 | cont | x |  |
| Serrano-Duenas 2005 | cont | x |  |
| Serrano-Duenas 2005 | cont | x |  |
| Serrano-Duenas 2005 | cont | x |  |
| Serrano-Duenas 2005 | cont | x |  |
| Serrano-Duenas 2005 | cont | x |  |
| Serrano-Duenas 2005 | cont | x |  |
| Serrano-Duenas 2005 | cont | x |  |
| Sauer 2017 | cont | x |  | Cognitive performance score |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Sauer 2017 | cont | x |  |
| Saczynski 2012 | cont | x |  | Cognitive performance score |
| Saczynski 2012 | cont | x |  |
| Saczynski 2012 | cont | x |  |
| Richardson 2021\_2 | cont | x |  | Cognitive performance score |
| Richardson 2021\_2 | cont | x |  |
| Racine 2020 | cont |  | x | Taken to “change” from baseline section *– The data is definitely change data, the first follow-up timepoint compares to baseline, thereafter from each follow-up timepoint (e.g. 2 months is compared to 1 month, 3 months to 2 etc). I think this is change data, rather than comparing scores at each timepoint.* |
| Racine 2020 | cont |  | x |
| Racine 2020 | cont |  | x |
| Pandharipande 2013 | cont | x |  | Cognitive performance score |
| Pandharipande 2013 | cont | x |  |
| Pandharipande 2013 | cont | x |  |
| Pandharipande 2013 | cont | x |  |
| Olofsson 2018 | cont | x |  | Cognitive performance score |
| Olofsson 2018 | cont | x |  |
| Olofsson 2018 | cont | x |  |
| Ojagbemi 2020 | cont | x |  | Cognitive performance score |
| Ojagbemi 2020 | cont | x |  |
| Ojagbemi 2020 | cont | x |  |
| Neufeld 2015 | cont | x |  | Cognitive performance score - note also some continuous change data |
| Neufeld 2015 | cont | x |  |
| Neufeld 2015 | cont | x |  |
| Neufeld 2015 | cont | x |  |
| Neufeld 2015 | cont | x |  |
| Neufeld 2015 | cont | x |  |
| Nerdal 2022 | cont | x |  | Cognitive performance score |
| Nerdal 2022 | cont | x |  |
| Nerdal 2022 | cont | x |  |
| Muller 2023 | cont | x |  | Is this a performance score? *Yes, I think this is just a total mean score at V3 (3 months)* |
| McCusker 2001 | cont | x |  | Cognitive performance score |
| McCusker 2001 | cont | x |  |
| McCusker 2001 | cont | x |  |
| McCusker 2001 | cont | x |  |
| Lingehall 2017 | cont | x |  | Cognitive performance score |
| Lingehall 2017 | cont | x |  |
| Lingehall 2017 | cont | x |  |
| Kunicki 2023 | cont | x |  | Cognitive performance score |
| Kunicki 2023 | cont | x |  |
| Kunicki 2023 | cont | x |  |
| Kunicki 2023 | cont | x |  |
| Kunicki 2023 | cont | x |  |
| Kunicki 2023 | cont | x |  |
| Kunicki 2023 | cont | x |  |
| Kunicki 2023 | cont | x |  |
| Kunicki 2023 | cont | x |  |
| Kunicki 2023 | cont | x |  |
| Kunicki 2023 | cont | x |  |
| Krogseth 2023 | cont | x |  | Cognitive performance score |
| Krogseth 2023 | cont | x |  |
| Krogseth 2016 | cont |  | x | Took to “change” from baseline section – *the data is presented as yearly decline from baseline, so a decline score/change score for our purpose* |
| Koster 2012 | cont | x |  | Cognitive performance score |
| Kat 2008 | cont | x |  | Cognitive performance score |
| Kat 2008 | cont | x |  |
| Kat 2008 | cont | x |  |
| Kat 2008 | cont | x |  |
| Kainz 2022 | cont | x |  | Cognitive performance score |
| Janssen 2021 | cont | x |  | Cognitive performance score |
| Janssen 2021 | cont | x |  |
| Jankowski 2011 | cont | x |  | Cognitive performance score |
| Jankowski 2011 | cont | x |  |
| Jankowski 2011 | cont | x |  |
| Jankowski 2011 | cont | x |  |
| Jankowski 2011 | cont | x |  |
| Inouye 2016 | cont | x |  | Cognitive performance score |
| Inouye 2016 | cont | x |  |
| Inouye 2016 | cont | x |  |
| Inouye 2016 | cont | x |  |
| Inouye 2016 | cont | x |  |
| Inouye 2016 | cont | x |  |
| Inouye 2016 | cont | x |  |
| Inouye 2016 | cont | x |  |
| Humbert 2021 | cont | x |  | Cognitive performance score |
| Humbert 2021 | cont | x |  |
| Hoogma 2023 | cont | x |  | Cognitive performance score |
| Giroux 2021 | cont | x |  | Cognitive performance score |
| Francis 1992 | cont | x |  | Cognitive performance score |
| Eide 2016 | cont | x |  | Cognitive performance score |
| Eide 2016 | cont | x |  |
| Duppils 2004 | cont |  | x | Is this a change from the baseline?? Check the result text section! – *Yes definitely a change, we used the values from Figure 1 to confirm the values were the follow-up score subtracted from baseline score. E.g. score in the Delirium group was 20.44 at follow-up, at baseline was 23.06, so -2.6 change, which matches what is reported.* |
| Cole 2008 | cont | x |  | Cognitive performance score |
| Cole 2008 | cont | x |  |
| Cole 2008 | cont | x |  |
| Chan 2017 | cont | x |  | Cognitive performance score |
| Chan 2017 | cont | x |  |
| Cavallari 2017 | cont | x |  | Cognitive performance score |
| Bulic 2020 | cont | x |  | Cognitive performance score |
| Bulic 2020 | cont | x |  |
| Bulic 2020 | cont | x |  |
| Bruck 2018 | cont | x |  | Cognitive performance score |
| Brown 2018 | cont | x |  | Cognitive performance score |
| Brown 2018 | cont | x |  |
| Cirbus 2019 | cont |  | x | Took to “change” from baseline section – *We had to use p value and sample size, so change assumption based on the statistics used, which generated a unstandardised Beta.* |
| Cirbus 2019 | cont |  | x |
| Cirbus 2019 | cont |  | x |
| Cirbus 2019 | cont |  | x |
| Cirbus 2019 | cont |  | x |
| Cirbus 2019 | cont |  | x |
| Ditzel 2023 | cont | x |  | Cognitive performance score |
| Sánchez-Lozano 2023 | cont | x |  | Cognitive performance score |
| Chen 2017 | cont | x |  | Cognitive performance score |

**Functional outcome**

**Categorical /dichotomous**

| Studies | Data type | functional change  **(decline vs no-decline)** | functional performance  **(impairment vs no-impairment)** | Note |
| --- | --- | --- | --- | --- |
| Abelha 2013 | cat |  | x | Impairment. It is ok. “The patients’ responses were categorized into two  groups: able or unable to perform each activity or group  of activities.” Become dependent in personal ADL&IADL *– Looks like performance, categories based on follow up survey response* |
| Abelha 2013 | cat |  | x |
| Alzoubi 2022 | cat |  | x | Impairment. It is ok. They classified as “dependent” vs “independent” in the table  “The maximum score is 6 indicating total independence, and the minimum score is 0 indicating total dependence, whereas a score between 1 and 5 indicates partial dependence.” *– Seems very clear* |
| Beishuizen 2020 | cat | x |  | it is ok. “Increase of at least one point on the 15-item modified Katz Index of Activities of Daily Living or Groningen Activity Restriction Scale at follow up compared with baseline” – *Seems very clear* |
| Bickel 2008 | cat |  | x | ??No description on whether it was below the threshold nor it was a change from the baseline.  Incident need for long-term care – agree to put in Performance, there is argument that could be a dichotomous measure of decline but seems to fit better as a cross-sectional follow-up assessment |
| Bickel 2008 | cat |  | x |
| Buurman 2011 | cat | x |  | it is ok. Functional decline was defined as a loss of at least one point on the original Katz ADL index score one year after hospital admission compared to the premorbid Katz ADL index score – *Clearly decline* |
| Czyzycki 2022 | cat |  | x | It is ok. “modified Rankin Scale was used to assess functional outcome. Unfavourable outcomes were defined as a modified Rankin Scale score of 3–6.”  Poor outcome - *Very clear Performance measure* |
| Czyzycki 2022 | cat |  | x |
| Czyzycki 2022 | cat |  | x |
| Czyzycki 2022 | cat |  | x |
| Czyzycki 2022 | cat |  | x |
| Czyzycki 2022 | cat |  | x |
| Davis 2012 | cat | x |  | it is ok. “Function (at least one category decline in five-point scale from independent to fully dependent for all care needs) between baseline and first follow-up in individuals also experiencing delirium” *– Categorical decline* |
| DelaVarga-Martínez 2022 | cat |  | x | ??The study presented functional outcome before and after surgery in delirium and no-delirium group. They used “independent and less mobility as a variable when compared in delirium and no-delirium group. No threshold to classify this condition reported. *As with cognitive outcomes, it does not look like they classified participants as a ‘decline’ vs ‘not decline’ based on change across time points, they rather use two x-sectional measures and looked at 2x2 data*. |
| DelaVarga-Martínez 2022 | cat |  | x |
| Durlach 2023 | cat |  | x | I think they classified into two categories. See Table 6. BADL and IADL cut of scores.  Only said “were classified according to BIADL” “Basic activities of daily living (BADL) three months after dis-charge were found to be preserved in 70.86% of the non-delirium group, 57.14% of the SSD group and 50% of the delirium group.”  BADL less preserved, IADL less preserved *– Looked at table 6, definitely performance categories based on lack of decline* |
| Durlach 2023 | cat |  | x |
| Edelstein 2004 | cat | x |  | I’m thinking functional performance but the study report as functional change  “Each of these basic and instrumental activities of daily living was rated on a scale of 0–4, with 0 being completely dependent and 4 being completely independent in that activity” “The patient’s ambulatory status was given a score between 1 (independent ambulator) and 6 (household ambulator with walker/crutches) for the prefracture status and 1 (independent ambulator) and 7 (restricted to wheelchair or bedridden) for post fracturestatus”*- Noted in the methods section, that it categorises scores based on return to prefracture level, or declined. Similarly, Table 3 uses the categories, which are based on change in score from baseline, so this seems to be in the Decline group.* |
| Edelstein 2004 | cat | x |  |
| Edelstein 2004 | cat | x |  |
| Eeles 2012 | cat |  | x | it is ok. “Although the FI can be considered as a continuum with higher values representing greater frailty, 0.25 has been proposed as the cut-off between ‘fit’ and ‘frail’” *– Measured at follow up and cut-off applied to categorise.* |
| FialhoSilva 2021 | cat |  | x | it is ok. “delirium as a predictor of functional outcome  (mRS>2).”… “Worse outcomes (mRS>2) at 30…..” *– Checked, categorical performance at follow up used.* |
| FialhoSilva 2021 | cat |  | x |
| Francis 1990 | cat | x |  | Not sure. “Nearly fourth of each group  reported some increase in dependency.” “Continuous  variables categorized using clinically meaningful cut points to generate odds ratio.” *– There is a lack of detail in the manuscript as to how exactly the ADLs were score, but seems that the difference between baseline and follow-up was the key outcome measure. “No significant differences were seen in rates of decline in ADL”* |
| Francis 1992 | cat |  | x | it is ok. “…..analyzed independent  community living as a dichotomous variable, with success  indicating that a subject was alive, not institutionalized,  and not dependent in basic ADL.”  Loss of independent community living *– Seems clear* |
| Furlaneto 2007 | cat | x |  | I’m thinking decline. “Functional and cognitive status, we used repeated  measures analysis of variance to compare the baseline and follow-up ADL, IADL, and BDRS of individual patients” Functional loss *- Agree, data seems to be a classification based on the change from baseline to follow up* |
| Gandossi 2023 | cat |  | x | It is ok. “ We defined the presence of poor functional status at 4 months with a total score < = 2.”  Poor functional status - *Seems very clear* |
| Gandossi 2023 | cat |  | x |
| Givens 2008 | cat | x |  | It is ok. “>=1-Point Decline in ADLs’”, “Loss of Prefracture Ability to  Walk 15 Feet Independently.” – *Based on methodological description and table 2, is clearly a decline measure as change from baseline was categorised.* |
| Givens 2008 | cat | x |  |
| Givens 2008 | cat | x |  |
| Givens 2008 | cat | x |  |
| Givens 2009 | cat | x |  | it is ok. “Increase of two or more ADL deficits at 1 month from baseline.” *– as above* |
| Guenther 2020 | cat | x |  | it is ok. “Patients’ demographic data were grouped into whether they had postoperative ADL lower than preoperative ADL (“ADL decline”), or whether there was no change or a higher ADL than before surgery (“No decline”).” *Seems very clear* |
| Hawley 2023 | cat |  | x | I think this impairment?? “Return to mobility, that is, at least mobile outdoors with aid(s) or frame among those mobile to this degree pre-fracture” “Mobility analysis conducted only among those with outdoor mobility (with or without aid(s)/frame) before hip fracture.”  Not Return to outdoor mobility – *This one is confusing, because there is no information about how return to own home and return to outdoor mobility were assessed, presumable just by interview. We think they have just calculated the number at 120 days and divided by the number at baseline and worked out percentage achieving the outcome…not sure, Hannah check.* |
| Hempenius 2016 | cat | x |  | It is ok. “ADL functioning, was categorized in a lower score at 3-month follow Up compared to the baseline the baseline score (“decreased”) versus a same or higher score (“same/ increased”). “Use of supportive care was dichotomized in an increased number of hours  supportive care per week at 3-month follow-up compared to baseline (“increased”) versus the  same or a decreased number of hours supportive care (“same/ decreased”).” - *Clearly described as a categorically defined change from baseline.* |
| Hempenius 2016 | cat | x |  |
| Hempenius 2016 | cat | x |  |
| Hempenius 2016 | cat | x |  |
| Hempenius 2016 | cat | x |  |
| Inouye 1998 | cat | x |  | it is ok. “At 3-month follow-up, ADL decline  was defined as a decline in ADL score from prehospitalization status until 3-month follow-up among survivors only”- Clear, decline |
| Jackson 2014 | cat |  | x | Check! I think this impairment “IADL disability (>8 on the FAQ) was seen in 26%  (108/422) and 23% (87/372) of individuals at 3 and 12 months”  “At least partial ADL disability (>0 on the Katz ADL), was seen in 32% (139/428) and 27%  (102/374) of individuals at 3 and 12 months”  “When comparing patients with and without baseline IADL disability, respectively, we found that 56% (19/34) vs. 23%  (87/384) had IADL disability at 3 months follow-up. Similarly at 12 months, 62% (21/34) vs. 20% (66/333) had IADL disability among those with and without baseline IADL disability – *This one is tricky, but the data taken from Table 4 seems to be Performance based, in that for each f-up time the two groups were compared and the OR calculated to determine the odds of the delirium group having a ‘worse’ or ‘better’ score* |
| Jackson 2014 | cat |  | x |
| Jackson 2014 | cat |  | x |
| Jackson 2014 | cat |  | x |
| Lee 2011 | cat |  | x | I think this impairment. Postoperative functional outcomes were evaluated  using activity levels, which were defined as  follows: I, normal; II, essentially independent outdoors  but requiring help with some activities; III, independent  indoors but always requiring help outdoors; IV, not  independent indoors but able to transfer and walk independently; and V, confined to bed or chair and not ambulatory *– No evidence of change from baseline, instead a timepoint cross-sectional assessment done.* |
| Lee 2011 | cat |  | x |
| Liang 2014 | cat | x |  | I think this is decline. “To compare functional changes at 1, 3,6, and 12 months, ADL or IADL functional decline was defined as lower ADL or IADL score at follow-up than at baseline” *- Data clearly displayed as categorical decline based on change in scores from baseline.* |
| Liang 2014 | cat | x |  |
| Liang 2014 | cat | x |  |
| Liang 2014 | cat | x |  |
| Liang 2014 | cat | x |  |
| Liang 2014 | cat | x |  |
| Liang 2014 | cat | x |  |
| Liang 2014 | cat | x |  |
| McCusker 2014 | cat | x |  | “10-point decline in Barthel Index.”*- correct, categorically defined decline from baseline.* |
| Miyamoto 2021 | cat |  | x | Impairment “BI score, which ranges between 0 (total dependence when performing ADL) and 100 (fully independent in performing ADL). There is no standardized cut-off value, but we applied a strict definition of disability affecting ADL as BI score <= 60” = ADL disability *– seems straightforward* |
| Miyamoto 2021 | cat |  | x |
| Morandi 2014 | cat |  | x | Impairment?? “The primary outcome was that of walking dependence captured as a trajectory from discharge to 1-year follow-up. Degree of walking dependence at discharge and at 1-year follow-up was assessed using the BI walking mobility subitem. A score less than 15 (the maximum score) is robust to the presence of mobility impairment”  Walking dependency – Used a clear cut-off, no evidence that they have subtracted baseline scores from follow-up, rather presented the percentages based on x-sectional results for each timepoint (discharge and follow-up). |
| Murray 1993 | cat | x |  | I think this is decline” Physical Function Outcome at Three Months After Discharge Compared to Admission Function”  “Second, we used a normative transformation of change  analysis in a linear regression model to examine the effect of incident delirium on change in function over the initial 3-month observation period.” *– From methods description, explains how they calculated differences between admission scores and 3-month scores, and then categorised these differences – so categorical decline.* |
| Murray 1993 | cat | x |  |
| Neufeld 2015 | cat |  | x | Impairment” Patients reporting falls since discharge”  Fall – *very clear as falls is a post-delirium measure – like mortality* |
| Neufeld 2015 | cat | x |  | Decline “Change in living status, baseline to 18-month” *– median change scores presented in results table 2, so this is decline* |
| Noriega 2015 | cat | x |  | Decline “Functional decline was defined as the loss of at least 1 point in the ADL total score with respect to the preadmission status.” – seems very clear |
| Noriega 2015 | cat | x |  |
| Noriega 2015 | cat | x |  |
| Noriega 2015 | cat | x |  |
| Ogawa 2017 | cat |  | x | Impairment” Frailty was defined as diminished handgrip strength (<26 kg for men and <18 kg for women) and/ or usual walking speed (<0.8 m/s)”  Frailty *- Seems very clear* |
| Paulino 2023 | cat |  | x | Impairment “Each function is scored as either independent (yes) or dependent (no). Based on their total score, the patients were categorized as follows: total dependency = 0; severe dependence = 1–2; moderate dependence = 3–4; mild dependence = 5; and total independence = 6.”  ADL dependency – *seems to be that that it is categorised based on performance from the follow up measure.* |
| Paulino 2023 | cat |  | x |
| Qu 2018 | cat |  | x | Impairment “we defined a poor outcome as an MRS score greater than to 3 or an IADL total score greater than 75th percentile of the IADL score.”  Poor outcome – *very clear, not a decline just based on follow up data* |
| Qu 2018 | cat |  | x |
| Qu 2018 | cat |  | x |
| Qu 2018 | cat |  | x |
| Quinlan 2011 | cat | x |  | Decline “Functional decline was defined as decline in at least one item from baseline. Those with preserved or improved function were considered not to have declined.” – very clear |
| Racine 2018 | cat | x |  | Decline “New impairment in cognitive IADL is impairment at 1 month not present at baseline in ability to use money, manage medications, use imbedded figure  relates to the telephone, or cooking.  Decline in physical function was defined as a decline of 5 or more points (0.5 population SD) on a composite physical functioning score based on ADLs, IADLs, and numbers in the columns.” – *Table 4 and methods clearly described the baseline to follow up change and the cut-offs used to define categories of change, Decline.* |
| Racine 2018 | cat | x |  |
| Rawle 2021 | cat |  | x | Impairment “The presence of delirium was instead strongly associated with the highest degrees of functional impairment, represented by a FAST score of >= 6” – Very clear, performance |
| Rollo 2022 | cat |  | x | Impairment “A score of 2 or less was considered the cutoff for functional independence”  Poor functional outcome *– Very clear* |
| Rudolph 2010 | cat | x |  | Decline “Functional decline was defined as the loss of 2 IADL points, which correlates to the decline of one IADL or partial decline on two IADLs” – *Very clearly defined decline* |
| Rudolph 2010 | cat | x |  |
| Shim 2015 | cat | x |  | Decline “Change in activities of daily living (preoperative to 30 days postoperatively” *Very clearly defined decline* |
| Shim 2015 | cat | x |  |
| Singler 2014 | cat |  | x | No clear information but I think it is impairment “IADL impairments were equally frequent  and the rate of patients with a history of  falls during the past 90 days was high”  Fall – *Falls outcomes generally have to be performance based, so performance is correct* |
| Suraarunsumrit 2022 | cat | x |  | Decline “A functional decline was defined as a  drop of 2 or more points from the preoperative IADL score to that achieved 3 months after surgery” Declined ADL – *very clear* |
| Suraarunsumrit 2022 | cat | x |  |
| Suraarunsumrit 2022 | cat | x |  |
| Tavares 2021 | cat | x |  | Decline “FD was defined as any decline in one or more points in the KI between the three moments at which the assessment was conducted.” Baseline, discharge, follow-up. *– Clear that it is decline, but note same limitation that could be between any of the timepoints.* |
| To-adithep 2023 | cat |  | x | Impairment. “B-ADL scores <=70 and I-ADL scores <9 were defined as dependency state” – *very clear* |
| To-adithep 2023 | cat |  | x |
| VanderHeijden 2023 | cat |  | x | Impairment “The total score ranges from 8 to 56, where fatigue is indicated by a score of 27 or higher”  Fatigue – *Very clear* |
| Verloo 2016 | cat |  | x | Impairment “the total score from 0-3 points indicates no frailty ….. In order to dichotomise between frail and not frail, a score of 6 points or more was considered as frail”  “IADLS A score of16 indicates that the patient is independent, as documented elsewhere” *– clearly impairment* |
| Verloo 2016 | cat |  | x |
| Vives-Borrás 2019 | cat | x |  | Decline “Loss of at least 5 points in the 6-month Barthel score.” *– very clear* |
| Zakriya 2004 | cat |  | x | Impairment “Need assistance with different ADL and IADL measures” -*Found in the methods section, it states that some measures, included frailty and ADLs were only assessed at follow-up, so is performance. No clear cut-offs provided to be certain that true impairment existed, so we have to assume this is the case, as the table states ‘needs assistance with’ which is equivalent to impairment.* |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |
| Zakriya 2004 | cat |  | x |

**Functional outcome**

**Continuous and change**

| Studies | Data type | Change from baseline | Score at follow-up | Note |
| --- | --- | --- | --- | --- |
| Cartei 2022 | cont |  | x | ??? check! I’m thinking “Score at follow-up” but there is information “ADL and IADL score changes over time were compared between groups with and without delirium estimating the time group interactions in 2 repeated-measures analysis of variance models.”  Disability in ADL- *ANOVA does not assess change over time, rather compares means at specific timepoints, so this is Scores, not change* |
| Chan 2017 | cont |  | x | Performance |
| Chan 2017 | cont |  | x |
| Chan 2017 | cont |  | x |
| Cirbus 2019 | cont | x |  | I think this is a change from the baseline “In patients with a pre-illness OARS ADL of 15 (poorer baseline function), the six-month OARS ADL was lowered by −2.9 points (95%CI: −0.3 to −5.6) and −4.3 points (95%CI: −4.4 to −8.5) in delirium  secondary to metabolic disturbances and organ  dysfunction, respectively.” *– Agree, the data in output table seems to be based on the change in scores.* |
| Cirbus 2019 | cont | x |  |
| Cirbus 2019 | cont | x |  |
| Cole 2008 | cont |  | x | Performance |
| Cole 2008 | cont |  | x |
| Cole 2008 | cont |  | x |
| Cole 2008 | cont |  | x |
| Cole 2008 | cont |  | x |
| Cole 2008 | cont |  | x |
| Decrane 2012 | cont |  | x | Performance  Fall |
| Decrane 2012 | cont |  | x |
| Decrane 2012 | cont |  | x |
| Eeles 2012 | cont |  | x | Performance  Frailty |
| Eide 2016 | cont |  | x | Performance and change result presented. Table 2 for performance (the mean difference between delirium and no-delirium group at follow-up timepoint) and Fig.1 is the change from the baseline. It was not clear which p value is the comparison between delirium and no-delirium group. Both delirium and no-delirium p-vale presented.  *The Figure only has within group changes and P value differences across visits. So must be performance data for the between group differences.* |
| Eide 2016 | cont |  | x |
| Eide 2016 | cont |  | x |
| Eide 2016 | cont |  | x |
| Eide 2016 | cont |  | x |
| Eide 2016 | cont |  | x |
| Eide 2016 | cont |  | x |
| Eide 2016 | cont |  | x |
| Fick 2013 | cont | x |  | Change |
| Giroux 2021 | cont |  | x | Performance |
| Giroux 2021 | cont | x |  | Change |
| Hoogma 2023 | cont |  | x | Performance |
| Hshieh 2017 | cont |  | x | Performance |
| Hshieh 2017 | cont |  | x |
| Hshieh 2017 | cont |  | x |
| Hshieh 2017 | cont |  | x |
| Humbert 2021 | cont |  | x | Performance |
| Humbert 2021 | cont |  | x |
| Isaia 2009 | cont | x |  | I’m thinking “change” “Similar results were found in ADL  functions: delirious patients lost more functions (2.8±  2.2) than those who did not develop delirium during  hospitalisation (0.7±0.6) (P<.001).” -  *There is not much information in the paper, so based on the above description from results, it seems they have subtracted scores from baseline for each group and compared the change and provided p value.* |
| Jankowski 2011 | cont |  | x | Performance |
| Jankowski 2011 | cont |  | x |
| Jankowski 2011 | cont | x |  | Change |
| Jankowski 2011 | cont | x |  |
| Katz 2001 | cont | x |  | Change |
| Liang 2014 | cont |  | x | Performance |
| Liang 2014 | cont |  | x |
| Liang 2014 | cont |  | x |
| Liang 2014 | cont |  | x |
| Liang 2014 | cont |  | x |
| Liang 2014 | cont |  | x |
| Liang 2014 | cont |  | x |
| Liang 2014 | cont |  | x |
| McCusker 2001 | cont |  | x | Performance |
| McCusker 2001 | cont |  | x |
| McCusker 2001 | cont |  | x |
| McCusker 2001 | cont |  | x |
| McCusker 2001 | cont |  | x |
| Miyamoto 2021 | cont |  | x | Performance |
| Miyamoto 2021 | cont |  | x |
| Monacelli 2018 | cont |  | x | Performance |
| Monacelli 2018 | cont |  | x |
| Monacelli 2018 | cont |  | x |
| Monacelli 2018 | cont |  | x |
| Murray 1993 | cont |  | x | Performance |
| Murray 1993 | cont |  | x |
| Murray 1993 | cont |  | x |
| Murray 1993 | cont |  | x |
| Neufeld 2015 | cont |  | x | Performance |
| Neufeld 2015 | cont |  | x |
| Neufeld 2015 | cont | x |  | Change |
| Neufeld 2015 | cont | x |  |
| Ojagbemi 2020 | cont | x |  | I’m thinking “change” “full delirium (MD=-5.6, 95% C.I = -9.0, -2.1, p = 0.002) at baseline independently predicted poorer performance in global cognitive functioning at 3 months poststroke. In the same table, significant declines in physical functioning (MD = -2.8, 95% C.I = -5.5,-0.2) were recorded at 3 months in stroke survivors with baseline DSM V delirium. However, these changes were not significant in those with ADS” – *It looks like the authors calculated mean change for each group from baseline and then compared using linear regression analysis.* |
| Oldenbeuving 2011 | cont |  | x | Performance |
| Pasinska 2019 | cont |  | x | Performance |
| Pasinska 2019 | cont |  | x |
| Pasinska 2019 | cont |  | x |
| Pasinska 2019 | cont |  | x |
| Rudolph 2010 | cont |  | x | Performance |
| Rudolph 2010 | cont |  | x |
| Sheng 2006 | cont |  | x | Performance |
| Sheng 2006 | cont |  | x |
| Sheng 2006 | cont |  | x |
| Shi 2019 | cont |  | x | Performance |
| Shi 2019 | cont |  | x |
| Shi 2019 | cont |  | x |
| Shi 2019 | cont |  | x |
| Shi 2019 | cont |  | x |
| Shi 2019 | cont |  | x |
| Shi 2019 | cont |  | x |
| Shi 2019 | cont |  | x |
| Shi 2019 | cont |  | x |
| Shi 2019 | cont |  | x |
| Shi 2019 | cont | x |  | Change |
| Shi 2019 | cont | x |  |
| Shi 2019 | cont | x |  |
| Shi 2019 | cont | x |  |
| Shi 2019 | cont | x |  |
| Shi 2019 | cont | x |  |
| Shi 2019 | cont | x |  |
| Shi 2019 | cont | x |  |
| Shi 2019 | cont | x |  |
| Shi 2019 | cont | x |  |
| Shi 2019\_2 | cont | x |  | Change “ADL decline was calculated by using follow-up ADLs score minus baseline ADLs score, which indicated the decline of activities of daily living” – seems clear that its decline. |
| Suraarunsumrit 2022 | cont |  | x | Performance |
| Suraarunsumrit 2022 | cont |  | x |
| Svenningsen 2014 | cont |  | x | Performance |
| Svenningsen 2014 | cont |  | x |
| VandenBoogaard 2012 | cont |  | x | Performance |
| Verloo 2016 | cont |  | x | Performance |
| Vida 2006 | cont | x |  | I’m thinking “change” but the study is not that clear  “In the non-dementia stratum, the group without delirium declined minimally, if at all, whereas the group with delirium declined to a significantly greater degree by 18 months” – I’m unsure but agree generally, it looks more like change data from the results. |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vida 2006 | cont | x |  |
| Vives-Borrás 2019 | cont |  | x | Performance |
| Vives-Borrás 2019 | cont |  | x |
| Wang 2021 | cont |  | x | Performance |
| Weng 2019 | cont |  | x | Performance |
| Weng 2019 | cont |  | x |
| Weng 2019 | cont |  | x |
| Whittamore 2014 | cont |  | x | Performance |
| Witlox 2013 | cont |  | x | Performance |
| Witlox 2013 | cont |  | x |
| Zakriya 2004 | cont |  | x | Performance |
| Zakriya 2004 | cont |  | x |
| Zipprich 2020 | cont |  | x | Performance |
| Miu 2013 | cont |  | x | Performance – ok, definitely cross sectional. |
| Miu 2013 | cont |  | x |
| Miu 2013 | cont |  | x |
| Miu 2013 | cont |  | x |
| Sánchez-Lozano 2023 | cont |  | x | Performance |