GRADE Protocol Handout

Step 1. Grade the starting quality of evidence for each guideline recommendation based on study designs (see GRADE Handbook for details): https://gdt.gradepro.org/app/handbook/handbook.html#h.fd20fe6dy7ea

- Grade the aggregate of studies relevant to each recommendation.
- The GRADE levels for the quality of the evidence are 'high', 'moderate', 'low' or 'very low'. GRADE's approach begins with the study design.
- If even a single **randomized trial** is available, the starting level is *'high quality'*; if no randomized trial is available but **observational studies** are available, the starting level is *'low quality'* (see table below).
- **In cases where a recommendation is based on expert opinion, the quality of evidence will be 'very low'** because it is based on anecdotal clinician observation (4).
- *When many outcomes are possible for a guideline recommendation, the grade for the overall quality of evidence is based on the grade for the outcome with the lowest quality of evidence, if that outcome is critical.*
- *Thus, critical outcomes determine the rating of quality of evidence across outcomes.*
- *Indicate reasons for upgrading or downgrading.*

Table 4: Overall quality of outcome evidence in GRADE

Level	Description
High	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low	Any estimate of effect is very uncertain

From *NICE Guidelines – Head Injury: Triage, Assessment, Investigations and Early Management*

Step 2: Consider the five factors that may reduce quality of RCTs from high to moderate, low or very low, and of observational studies from 'low' to 'very low' (for details, see https://gdt.gradepro.org/app/handbook/handbook.html#h.6wuuudis64a8 section 7.5.2)

(i) Risk of bias – **RoB rating in Excel document**

- **Low risk of bias –** no downgrade
- **Unclear risk of bias –** downgrade only if potential limitations are LIKELY to lower the confidence in the estimate of effect
- **High risk of bias –** downgrade 1 if crucial limitation in 1 criterion or some limitation in multiple criteria sufficient to reduce the confidence in the estimate of effect, or downgrade 2 if crucial limitation in multiple criteria, that would substantially reduce the confidence in the estimate of effect

(ii) Inconsistencies between studies

Reasons to downgrade

- **Wide variance of point estimates across studies** (note: direction of effect is not a criterion for inconsistency)

- **Minimal or no overlap of confidence intervals (CI)**, which suggests variation is more than what one would expect by chance alone
- **Statistical criteria, including tests of heterogeneity** which test the null hypothesis that all studies have the same underlying magnitude of effect, have a low p-value ($p < 0.05$), indicating to reject the null hypothesis
- See https://gdt.gradepro.org/app/handbook/handbook.html#h.g2dqzi9je57e **examples 1, 2 , 3**

(iii) Indirectness of evidence - Direct evidence consists of research that directly compares the interventions which we are interested in, delivered to the populations in which we are interested, and measures the outcomes important to patients (aka applicability)

- **Differences in population –** does the research use a sample we are interested in?
- **Differences in intervention/comparisons –** are we interested in the interventions and comparisons?
- **Differences in outcome -** study uses of substitute or surrogate outcomes rather than patient important outcomes

(iv) Imprecision in estimates – see table below

(v) A high probability of publication bias (ex. studies funded by for-profit industry are more likely to have bias).

Table 2: Description of quality elements in GRADE for intervention and diagnostic studies (adapted from quality elements for intervention studies)

Quality element	Description
Limitations	Intervention - Limitations in the study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect. Diagnostic - Cross sectional or cohort studies in patients with diagnostic uncertainty and direct comparison of test results with an appropriate reference standard are considered high quality. See also QUADAS-2 quality assessment checklist.
Inconsistency	Intervention - Inconsistency refers to an unexplained heterogeneity of results. Diagnostic - Unexplained inconsistency in sensitivity, specificity, or likelihood ratios (rather than relative risk or mean differences) can reduce quality of studies.
Indirectness	Intervention - Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question, or recommendation made. Diagnostic - Quality can be reduced if • important differences exist between populations studied and those for whom the recommendation is intended (in previous testing, spectrum of disease or comorbidity). • important differences exist in test studied and diagnostic expertise of people applying them in studies compared with settings for which recommendations are intended. • tests being compared are compared to a reference standard in different studies and not directly compared in the same studies.
Imprecision	Intervention - Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of the effect relative to the clinically important threshold. Diagnostic - Wide confidence intervals for estimates of test accuracy or true and false positive and negative rates can reduce quality of evidence.
Publication bias	Intervention - Publication bias is a systematic underestimate or an overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. Diagnostic - High risk publication bias (for example from small studies for new intervention or test, or asymmetry in funnel plot) can lower quality of evidence.

Source: Adapted from BMJ 2008 diagnostic GRADE paper,[230] GRADE working group.[230]

Summary *Table from NICE Guidelines – Head Injury: Triage, Assessment, Investigations and Early Management*

Step 3: Consider the three factors that may raise grading of observational studies from 'low' to 'moderate' or 'high':

- What evidence can be upgraded? Primarily observational studies that have low risk of bias, RCTs that have risk of bias that is not serious, and indirect RCTs (see indirectness) that are really convincing. Consider the three factors listed below when these conditions are met.

Note: Consideration of all the criteria for downgrading the estimate of the quality of evidence must precede consideration of reasons for upgrading the estimate quality.

(i) Large (+1) or very large (+2) and consistent estimates of a treatment effect;

a. see https://gdt.gradepro.org/app/handbook/handbook.html#h.6wuuudis64a8 (5.3.1)
b. Decisions to rate up quality of evidence because of large or very large effects (Table 5.9) should consider not only the point estimate but also the precision (width of the CI) around that effect: one should rarely and very cautiously rate up quality of evidence because of apparent large effects, if the CI overlaps substantially with effects smaller than the chosen threshold of clinical importance.

(ii) The presence of a dose–response gradient;

a. see https://gdt.gradepro.org/app/handbook/handbook.html#h.5zbbwi81tho4 (5.3.2) – **Examples 1 and 2**

(iii) A situation in which confounding is expected to reduce the magnitude of the effect or if confounding would increase effect but no effect was observed

a. When confounding is expected to reduce a demonstrated effect/controlling for covariates would probably increase effect size in these studies (Upgraded by One Level): See https://gdt.gradepro.org/app/handbook/handbook.html#h.6wuuudis64a8 (5.3.3) E**xample 2**

b. When confounding is expected to increase the effect but no effect was observed (Upgraded by One Level) See https://gdt.gradepro.org/app/handbook/handbook.html#h.6wuuudis64a8 (5.3.3) **Examples 3**

Table 2 The process of grading quality of evidence according to GRADE

Starting level based on study design	Reduce grade	Raise grade	Final level
Randomized trials = high	*Risk of bias* −1 level if serious −2 levels if very serious *Inconsistency* −1 level if serious −2 levels if very serious *Indirectness* −1 level if serious −2 levels if very serious *Imprecision* −1 level if serious −2 levels if very serious *Publication bias* −1 level if likely −1 levels if very likely	*Large effect* +1 level if large +2 levels if very large *Dose response* +1 level if evidence of a gradient *All plausible residual confounding* +1 level if would reduce a demonstrated effect +1 level if would suggest a spurious effect if no effect was observed	High Moderate
Observational studies = low ⇨			⇨ Low Very Low

Step 4. Determine the direction and strength of a recommendation.

- (a) Quality of evidence is only one of the four key factors determining the strength of a recommendation, according to GRADE (see table 3 below). The others are the (b) magnitude of the difference between the desirable and undesirable consequences, (c) the certainty about values and preferences of patients, and (d) the resource expenditure associated with the compared management options.
- Direction of a recommendation is either 'for' or 'against'. A recommendation is graded either 'strong' (i.e. 'We recommend…' for a positive recommendation or 'We do not recommend …' for a negative recommendation) or 'weak' ('We suggest …' or 'We do not suggest …'). On occasion, to avoid making statements about what should not be done (e.g. 'we recommend that treatment A is not used'), they may recommend an alternative option stating what should be done (e.g. 'we recommend that treatment B is used rather than treatment A').
- **In cases where a recommendation is based on expert opinion, the strength of recommendation will be 'strong'** because it is very likely that it is based on a cost/benefit analysis (3).

Table 3 Determinants of strength of recommendation, according to GRADE

Factor	Comment
Quality of evidence	Strong recommendations usually require higher quality evidence for all the critical outcomes. The lower the quality of evidence, the less likely is a strong recommendation.
Balance between desirable and undesirable effects	Panellists should make stronger recommendations for interventions that influence outcomes with high patient importance. If the baseline risk is different among different populations, they should made separate recommendations. The larger the difference between the desirable and undesirable effects, the higher the likelihood that a strong recommendation is warranted.
Values and preferences	The more values and preferences vary, or the greater the uncertainty in values and preferences, the higher the likelihood that a weak recommendation is warranted.
Costs (resource allocation)	The higher the incremental cost, all else being equal, the less likely that the recommendation in favor of an intervention is strong.

- From EFNS (1)

<u>Adapted from</u>

1. ***EFNS "Guidance for the preparation of neurological management guidelines by EFNS scientific task forces – revised recommendations 2012"***
- ***M. A. Leonea, M. Braininb, P. Boonc, M. Pugliattid, M. Keindlb and C. L. Bassett***
2. ***NICE Guidelines – "Head Injury: Triage Assessment, Investigations and Early Management"***
3. ***Brożek JL, Akl EA, Compalati E, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines part 3 of 3. The GRADE approach to developing recommendations. Allergy 2011;66:588-595.***
4. ***Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401-406.***