GRADE Training
BHIVA guidelines

Dr Ian Williams

March 2017

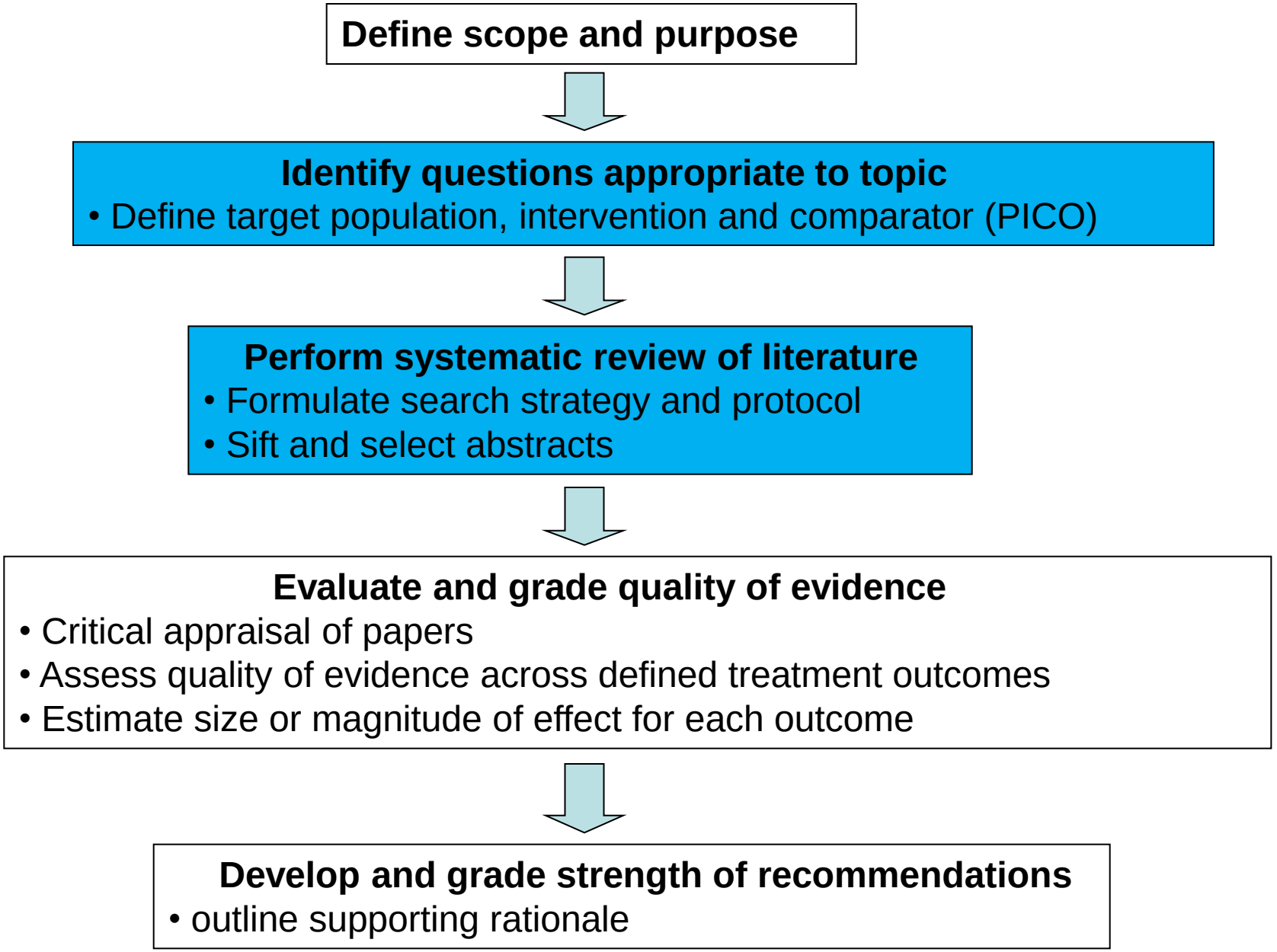
GRADE System:
Grades of recommendation, assessment, development and evaluation

BHIVA: Guideline development

British HIV Association (BHIVA) Guideline Development Manual

CONTENTS

http://www.bhiva.org/GuidelineDevelopmentManual.aspx


Define scope and purpose
Identify questions appropriate to topic
• Define target population, intervention and comparator (PICO)
Perform systematic review of literature
• Formulate search strategy and protocol
• Sift and select abstracts
Evaluate and grade quality of evidence
• Critical appraisal of papers
• Assess quality of evidence across defined treatment outcomes
• Estimate size or magnitude of effect for each outcome
Develop and grade strength of recommendations
• outline supporting rationale

PICO Framework

Select topics and define questions appropriate for topics

For each question define PICO criteria for literature search

- Population
- Intervention
- Comparator
- Outcome

For GRADE, it is vital to define outcomes for quality assessment of evidence and inform recommendation

Patient outcomes

For GRADE it is important to:

- Define patient outcomes
- Rank outcomes as:
  - critically important for decision making
  - important but not critical for decision making
  - not important for decision making

Assessment of evidence:

- Generate an estimate of effect for each outcome

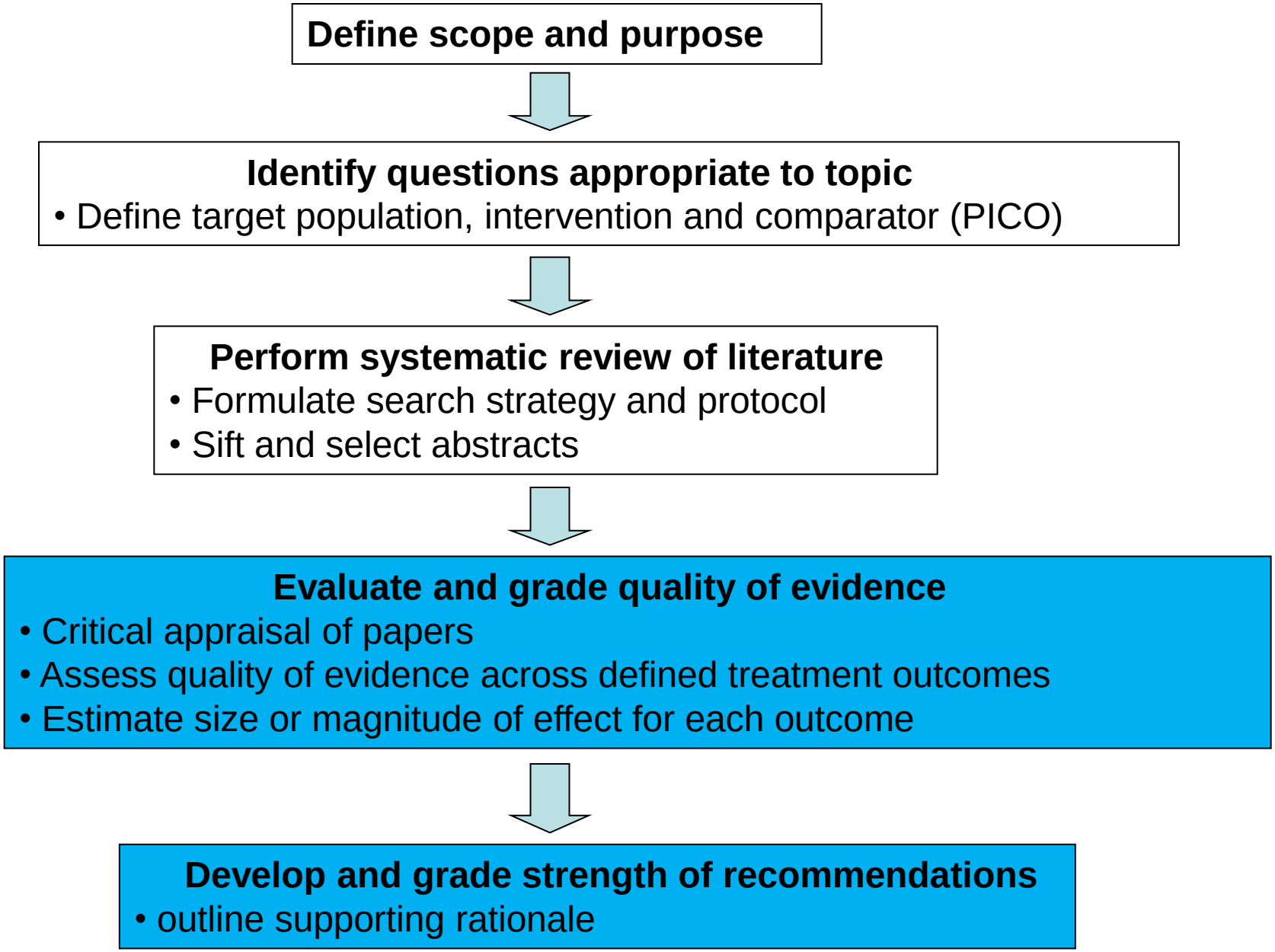
Literature search

Formulate search strategy and protocol

For literature search define:

- Data bases
- Date parameters
- Study design
- Conference abstracts

Following search sift and select studies which meet selection criteria.


Define scope and purpose
Identify questions appropriate to topic
• Define target population, intervention and comparator (PICO)
Perform systematic review of literature
• Formulate search strategy and protocol
• Sift and select abstracts
Evaluate and grade quality of evidence
• Critical appraisal of papers
• Assess quality of evidence across defined treatment outcomes
• Estimate size or magnitude of effect for each outcome
Develop and grade strength of recommendations
• outline supporting rationale

GRADE

‘Grading of Recommendations Assessment Development and Evaluation (GRADE) is an approach to grading evidence that moves away from initial reliance on study design to consider overall quality of evidence across outcomes’

GRADE System

Grading a recommendation: Two components

1. **Quality of evidence**:
   - extent to which confidence in estimate of effect adequate to support decision
   - High, moderate, very low, low
2. **Strength of recommendation**
   - strong or weak (conditional)

Recommendations should be specific and actionable concerning a target population and a specific intervention/strategy

Quality of evidence

Critical appraisal of papers
Estimate size or magnitude of effect for each outcome (forest plots)
Rate the quality of evidence for each outcome across studies

Rating is modified downward

- Study limitations
- Imprecision
- Inconsistency of results
- Indirectness of evidence
- Publication bias

Rating is modified upward

- Large magnitude of effect
- Dose response confounders likely minimise the effect

RCTs start with high rating, observational studies with a low rating
Final rating of quality for each outcome: high, moderate, low, very low

Quality of Evidence

Evidence and Summary of findings tables

- Provide details of evidence for each outcome
- Provide an estimate of effect for each outcome
- Assess if size of effect is clinically important / relevant
- Grade quality of evidence for each outcome (A-D)
- Grade importance of each outcome

Summary of Findings tables: PI monotherapy (2012 guidelines)

Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	PI monotherapy versus combination therapy	control	Relative (95% CI)	Absolute		
Virological suppresion (follow-up 48-96 weeks; viral load <50)												
10	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	476/591 (80.5%)	523/607 (86.2%)	RR 0.95 (0.9 to 0.99)	43 fewer per 1000 (from 9 fewer to 86 fewer)	⊕⊕⊕O MODERATE	CRITICAL
								87.3%		44 fewer per 1000 (from 9 fewer to 87 fewer)		

Strength of recommendation

Strong (1)

Weak or conditional (2)

Determined by:

- Quality of evidence
- Balance of desirable/undesirable outcomes
- Values and preferences
- Resource use

Can be 2 directional ie for or against a strategy

Strength of recommendation

- **Strong (1):** ‘we recommend’

  Implies that most patients and clinicians should follow this course of action but a small proportion may not if there is a good rationale not to.

- **Weak or conditional (2):** ‘we suggest’

  Implies that many patients and clinicians would want to follow this this strategy but many would not ie an alternative strategy may be reasonable depending on the patients circumstances and wishes.

Strength of recommendation

British HIV Association BHIVA

5.4 Which third agent

5.4.1 Recommendations

- We recommend therapy-naïve individuals start combination ART containing atazanavir/r, darunavir/r, dolutegravir, elvitegravir/c, raltegravir or rilpivirine as the third agent (1A).
- We suggest that for therapy-naïve individuals, efavirenz is an acceptable alternative third agent (1A).

? Is there an error in the GRADE recommendation

Strength of recommendation

Strong (1)

Weak or conditional (2)

Determined by:

- Quality of evidence
- Balance of desirable/undesirable outcomes
- Values and preferences
- Resource use

Can be 2 directional ie for or against a strategy

Modified GRADE system

Appendix 7

Summary of the modified GRADE system (grades 1A–2D)

1A

Strong recommendation.

High-quality evidence.

Benefits clearly outweigh risk and burdens, or vice versa.

Consistent evidence from well performed randomised, controlled trials or overwhelming evidence of some other form. Further research is unlikely to change our confidence in the estimate of benefit and risk.

Strong recommendations, can apply to most patients in most circumstances without reservation.

Clinicians should follow a strong recommendation unless there is a clear rationale for an alternative approach.

Modified GRADE system

2D

Weak recommendation.

Very low-quality evidence.

Uncertainty in the estimates of benefits, risks, and burdens; benefits may be closely balanced with risks and burdens.

Evidence limited to case studies and expert judgment.

Very weak recommendation; other alternatives may be equally reasonable.

Example: PI monotherapy in ART experienced patients

Question

- Is PI monotherapy an appropriate treatment strategy for treatment experienced patients on ART with virological suppression?

PI monotherapy

Search protocol

Population: ART experienced, >6 months VL <50, no previous PI resistance

Intervention: PI/rit monotherapy

Comparator: Standard triple HAART

Systematic reviews and RCTs

Search period: 1st January 2008 -16th September 2011

Data bases: Medline, Embase, Cochrane

Conference abstracts: 2009-2011

PI monotherapy

Switch/simplification/stopping

questions 10-12:

Medline: 375

Embase: 465

Cochrane: 168

Total (duplicates excluded): 489

Sifted and selected for PI monotherapy question:

- 18 papers identified for 10 studies
  (8:Lopinavir/r; 2:Darunavir/r)

PI Monotherapy

Treatment outcomes:

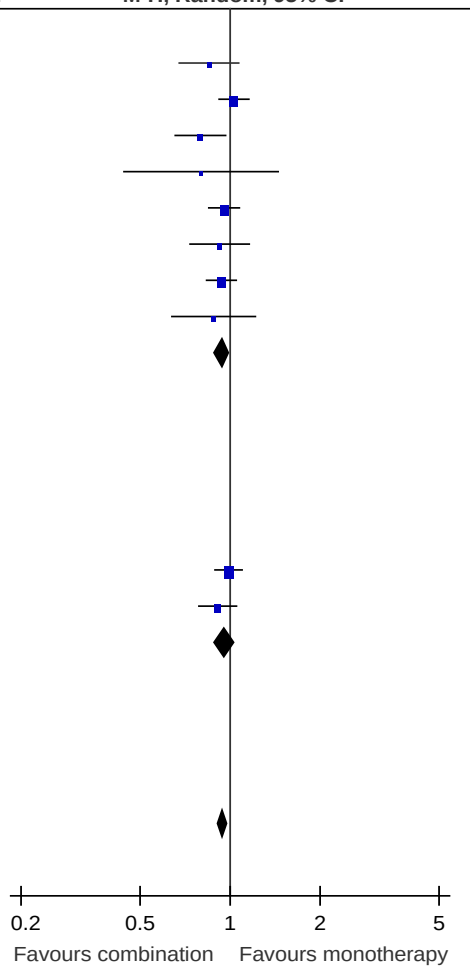
1. Virological suppression VL<50 at 48 +/- 96 weeks
2. HIV drug resistance
3. CD4 count increase
4. Serious adverse events
5. Grade 3/4 clinical events
6. Grade 3/4 laboratory events
7. Grade 3/4 abnormal LFTs
8. Grade 3/4 CNS disease

Ranked critical, important, not important

Generate an estimate of effect of the intervention for each outcome

Forest plot: PI monotherapy v combination therapy outcome: virological suppression

Study or Subgroup	PI monotherapy Events	PI monotherapy Total	Combination therapy Events	Combination therapy Total	Weight	Risk Ratio M-H, Random, 95% CI
1.1.1 Lopinavir						
Arribas 2005 (OK Pilot)	17	21	20	21	4.3%	0.85 [0.68, 1.07]
Cahn 2011 (wk51.4)	39	41	36	39	17.3%	1.03 [0.92, 1.16]
Gutmann 2010 (MOST)	23	29	31	31	6.0%	0.80 [0.66, 0.97]
Hasson 2011 (KAMON 2)	8	15	10	15	0.6%	0.80 [0.44, 1.45]
Meynard 2010 (KALESOLO)	73	87	87	99	16.3%	0.95 [0.85, 1.07]
Nunes 2009 (KalMo wk 96)	24	30	26	30	4.3%	0.92 [0.74, 1.16]
Pulido 2008 (OK04 wk48)	85	103	90	102	17.4%	0.94 [0.83, 1.05]
Waters 2008 (wk48)	18	26	22	28	2.2%	0.88 [0.64, 1.21]
Subtotal (95% CI)		**352**		**365**	**68.4%**	**0.94 [0.89, 1.00]**
Total events	287		322			
Heterogeneity: Tau² = 0.00; Chi² = 6.99, df = 7 (P = 0.43); I² = 0%						
Test for overall effect: Z = 2.12 (P = 0.03)						
1.1.2 Darunavir						
Arribas 2010 (MONET wk48)	107	127	110	129	20.8%	0.99 [0.89, 1.10]
Katlama 2010 (MONOI)	82	112	91	113	10.8%	0.91 [0.79, 1.05]
Subtotal (95% CI)		**239**		**242**	**31.6%**	**0.96 [0.88, 1.04]**
Total events	189		201			
Heterogeneity: Tau² = 0.00; Chi² = 0.88, df = 1 (P = 0.35); I² = 0%						
Test for overall effect: Z = 0.94 (P = 0.35)						
Total (95% CI)		**591**		**607**	**100.0%**	**0.95 [0.90, 0.99]**
Total events	476		523			
Heterogeneity: Tau² = 0.00; Chi² = 7.91, df = 9 (P = 0.54); I² = 0%						
Test for overall effect: Z = 2.28 (P = 0.02)						
Test for subgroup differences: Not applicable						



PI monotherapy: GRADE tables

Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	PI monotherapy versus combination therapy	control	Relative (95% CI)	Absolute		
Virological suppresion (follow-up 48-96 weeks; viral load <50)												
10	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	476/591 (80.5%)	523/607 (86.2%)	RR 0.95 (0.9 to 0.99)	43 fewer per 1000 (from 9 fewer to 86 fewer)	⊕⊕⊕O MODERATE	CRITICAL
								87.3%		44 fewer per 1000 (from 9 fewer to 87 fewer)		

Virological suppression: VL <50 48-96 weeks

PI Monotherapy:

476/519 (80.5%)

805 out of every 1000

Standard therapy

523/607 (86.2%)

862 out of every 1000

Relative risk 0.95 (95% CI 0.90-0.99)
ie 5% lower risk of virological suppression
True effect between 1-10 % lower risk

Absolute effect:
43 fewer people per 1000
Will maintain virological suppression

(from 9 fewer to 86 fewer per 1000)

PI monotherapy: GRADE tables

Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	PI monotherapy versus combination therapy	control	Relative (95% CI)	Absolute		
Virological suppresion (follow-up 48-96 weeks; viral load <50)												
10	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	476/591 (80.5%)	523/607 (86.2%) 87.3%	RR 0.95 (0.9 to 0.99)	43 fewer per 1000 (from 9 fewer to 86 fewer) 44 fewer per 1000 (from 9 fewer to 87 fewer)	⊕⊕⊕O MODERATE	CRITICAL
Drug resistance (follow-up 48 weeks; genotypic testing)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	2/125 (1.6%)	1/115 (0.9%) 3.9%	RR 1.15 (0.15 to 9.01)	1 more per 1000 (from 7 fewer to 70 more) 6 more per 1000 (from 33 fewer to 312 more)	⊕OOO VERY LOW	CRITICAL
Serious adverse events (follow-up 48-96 weeks; monitoring)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	28/280 (10%)	27/281 (9.6%) 9.7%	RR 1.05 (0.63 to 1.73)	5 more per 1000 (from 36 fewer to 70 more) 5 more per 1000 (from 36 fewer to 71 more)	⊕⊕⊕O MODERATE	IMPORTANT

No differences in all other outcomes but quality of evidence very weak/weak

BHIVA ART Guidelines 2012

‘We recommend continuing standard combination ART as the maintenance strategy in virologically suppressed patients (1C)’

BHIVA ART Guidelines

5.3 Which nucleoside reverse transcriptase inhibitor backbone

5.3.1 Recommendations

- We recommend therapy-naïve individuals start combination ART containing tenofovir-DF and emtricitabine or tenofovir-AF and emtricitabine as the preferred NRTI backbone (1A).
- We suggest abacavir and lamivudine is an acceptable alternative NRTI backbone in therapy-naïve individuals. In those with a baseline viral load >100,000, it should be used with caution if there are clinical reasons to prefer it over alternative NRTI backbones (2A).

The caution regarding baseline viral load does not apply if abacavir/lamivudine is used with dolutegravir (2A).

What is the basis for these recommendations, is there an error ?

BHIVA Guidelines

1.2.4 Good practice points (GPP)

- GPPs are recommendations based on the clinical judgment and experience of the working group. GPPs emphasise an area of important clinical practice for which there is not, nor is there likely to be, any significant research evidence. They address an aspect of treatment and care that is regarded as such sound clinical practice that health care professionals are unlikely to question it and where the alternative recommendation is deemed unacceptable.

Patient involvement

3.1 Recommendation:

- We recommend patients are given the opportunity to be involved in making decisions about their treatment. (GPP)

Educational resources

- BMJ series 2008: The BMJ published a series of 5 articles introducing the GRADE system and explaining how it works. It was aimed at clinicians and guideline writers, The articles can be accessed through the grade working group web site at : http://www.gradeworkinggroup.org

- McMaster GRADE on line modules: Includes 2 modules developed by grade working group for the WHO. The web address is: http://cebgrade.mcmaster.ca/