# Annex 8. GRADE evidence summary tables

Author(s): STREAM Stage 1 Trial investigators reported for the Guideline Development Group for the WHO treatment guidelines on MDR/RR-TB, 2018 update (6 July 2018) - FINAL RESULTS

Question: PICO 1. In MDR-TB patients, are treatment regimens lasting up to 12 months as likely to lead to the outcomes listed below when compared with those recommended in the WHO guidelines of 2011? (Cured/completed, culture conversion by 6 months, failure, relapse, survival (or death), adverse reactions, acquisition (amplification) of additional drug resistance, adherence to treatment interruption due to non-adherence)) [MAIN TABLE]

Setting: Assessment of final results from a phase III, multi-centre, international, parallel group, open-label, randomised controlled trial of a standardised MDR-TB treatment regimen lasting 9-11 months versus a longer regimen using a non-inferiority design. The trial enrolled patients between July 2012 and June 2015 in seven sites located in Ethiopia, Mongolia, South Africa and Viet Nam (intention to treat population = 424 [282 in study arm; 142 control arm]; mITT population = 369 [245 in study arm; 124 control arm]). Treatment allocation was not blinded to the participants, care givers or data managers. All local and reference laboratory assessments, including microbiological tests involved in the assignment of patient outcome, were conducted blind.

Bibliography: 1. ISRCTN - ISRCTN78372190: The evaluation of a standardised treatment regimen of anti-tuberculosis drugs for patients with multi-drug-resistant tuberculosis (MDR-TB). Available from: http://www.isrctn.com/ISRCTN78372190 2. Nunn AJ, Rusen ID, Van Deun A, Torrea G, Phillips PPJ, Chiang C-Y, et al. Evaluation of a standardized treatment regimen of anti-tuberculosis drugs for patients with multi-drug-resistant tuberculosis (STREAM): study protocol for a randomized controlled trial. Trials. 2014;15:353 3. Nunn A. Preliminary efficacy results of STREAM stage 1. In: STREAM trial (Evaluation of a Standardised Treatment Regimen of Anti-tuberculosis Drugs for Patients with Multidrug-resistant Tuberculosis): preliminary stage 1 results. Guadalajara, Mexico: UNION; 2017. Available from: http://www.professionalabstracts.com/theunion2017/ipplanner/#/person/3892

			Certainty a	assessment			№ of p	atients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a standardised 9 month shorter MDR-TB regimen	longer MDR-TB regimens	Relative (95% CI)	Absolute (95% CI)		
Time to cultu	re conversion by w	reek 20 (assessed wi	th: STREAM Stage 11	trial data (mITT popul	ation)) <sup>a</sup>							
1	randomised trials	not serious <sup>b</sup>	not serious	not serious °	serious <sup>d</sup>	none	е	f	HR 1.13 (0.91 to 1.40)	per 100 (from to)	⊕⊕⊕ MODERATE	IMPORTANT
Favourable o	outcome (follow up:	mean 132 weeks; as	ssessed with: STREAM	M Stage 1 trial data (n	nITT population)) <sup>9</sup>	!			ļ.			
1	randomised trials	not serious	not serious	not serious	serious <sup>d</sup>	none	193/245 (78.8%)	99/124 (79.8%)	RR 0.99 (0.88 to 1.01) <sup>NJ</sup>	8 fewer per 1,000 (from 8 more to 96 fewer)	MODERATE	CRITICAL
Favourable o	outcome (follow up:	132; assessed with:	STREAM Stage 1 tria	I data (per protocol)) <sup>9</sup>								
1	randomised trials	not serious	not serious	not serious	serious	none	186/227 (81.9%)	67/83 (80.7%)	RR 1.02 (0.90 to 1.15) <sup>h,j</sup>	16 more per 1,000 (from 81 fewer to 121 more)	⊕⊕⊕ MODERATE	CRITICAL
Died from an	I ny cause during trea	Latment or follow-up, a	Il cases (follow up: me	Lean 132 weeks; asses	Lesed with: STREAM S	Lage 1 trial data (ITT safety popul	ation)) <sup>kl</sup>					
1	randomised trials	not serious	not serious	not serious	very serious <sup>m</sup>	none	24/282 (8.5%)	9/141 (6.4%)	HR 1.38 (0.64 to 2.96) <sup>n</sup>	23 more per 1,000 (from 22 fewer to 114 more)	⊕⊕⊖⊖ Low	CRITICAL
Died from an	y cause during trea	atment or follow-up, p	eople living with HIV (	follow up: mean 132	weeks; assessed with	n: STREAM Stage 1 trial data (ITT	safety population))°					
1	randomised trials	not serious	not serious	not serious	very serious <sup>m</sup>	none	18/103 (17.5%)	4/50 (8.0%)	HR 2.23 (0.76 to 6.60) <sup>n</sup>	90 more per 1,000 (from 19 fewer to 343 more)	ФФС Low	CRITICAL

			Certainty a	assessment			N₂ of p	atients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a standardised 9 month shorter MDR-TB regimen	longer MDR-TB regimens	Relative (95% CI)	Absolute (95% CI)		
Lack of conv	ersion, reversion o	r relapse (follow up: n	nean 132 weeks; asse	essed with: STREAM	Stage 1 trial data (ml	TT population)) <sup>kp</sup>						
1	randomised trials	not serious	not serious	not serious	serious	none	26/245 (10.6%)	7/124 (5.6%)	RR 1.88 (0.84 to 4.21) h	50 more per 1,000 (from 9 fewer to 181 more)	MODERATE	CRITICAL
Lack of conv	ersion, reversion o	r relapse (follow up: 1	.32; assessed with: S'	FREAM Stage 1 trial o	data (per protocol)) <sup>k</sup>							
1	randomised trials	not serious	not serious	not serious	serious	none	24/227 (10.6%)	7/83 (8.4%)	RR 1.25 (0.56 to 2.80) <sup>h</sup>	21 more per 1,000 (from 37 fewer to 152 more)	MODERATE	CRITICAL
New onset Q	TcF interval prolon	gation to 500ms or m	ore on electrocardiog	ram (automated meas	surement) (assessed	with: STREAM Stage 1 trial data (	ITT safety population)) <sup>q</sup>					
1	randomised trials	not serious	not serious	not serious <sup>r</sup>	serious	none	26/282 (9.2%)	6/141 (4.3%)	RR 2.17 (0.91 to 5.14) h	<b>50 more per 1,000</b> (from 4 fewer to 176 more)	MODERATE	CRITICAL
Adverse ever	nt of grade 3 to 5 s	everity (follow up: me	an 132 weeks; asses	sed with: STREAM St	tage 1 trial data (ITT s	rafety population)) <sup>s</sup>						
1	randomised trials	not serious	not serious <sup>t</sup>	not serious	serious	none	136/282 (48.2%)	64/141 (45.4%)	RR 1.06 (0.85 to 1.32) hu	27 more per 1,000 (from 68 fewer to 145 more)	MODERATE	CRITICAL

CI: Confidence interval; HR: Hazard Ratio; RR: Risk ratio

- a. As per report to WHO by STREAM Stage I trial investigators on 6 July 2018
- b. The risk of bias was not downgraded for lack of blinding of patients and clinicians because this was an inherent study design and measures were taken to mitigate the blinding; cross-over from shorter to a longer regimen was also very low. No information provided about selection bias upstream of randomization
- c. The patient group is expected to be similar to the one who would receive the regimen but there was insufficient information about risk groups at the time of the review in January 2018 (e.g. frequency of cavitary disease on radiography, disease severity, exposure to second-line treatment, ARV in PLHIV). No trial data in children. Overall the trial eligibility criteria are similar to the ones recommended by WHO in the guidelines and there was no restriction in CD4 count among the PLHIV in the trial (one third of study participants). The main concern would be to extrapolate to patients presumed to have strains susceptible to fluoroquinolones and 2nd line injectable agents but in whom resistance has not been reliably excluded by DST.
- d. Single trial, number of patients in each arm is relatively small with wide confidence limits.
- e. Kaplan-Meier estimate of culture conversion at Week 20: 99.2% (95% CL: 95.8%, 99.9%)
- f. Kaplan-Meier estimate of culture conversion at Week 20: 99.2% (95% CL: 97.3%, 99.8%)
- g. Favourable outcome is defined as culture negative at 132 weeks and at the previous occasion that the patient was seen, unless the patient outcome had already been classified as unfavourable (see below). The mapping of trial endpoints to the WHO TB outcomes definitions (2005 or 2013) was not direct (e.g. some cases normally classified as Cured could be reclassified as unfavourable in the trial should death or relapse happen during the follow-up period).

- h. Unadjusted value
- i. The difference in proportions (1.1% (-7.7%, 9.8%)) becomes smaller after adjustment for HIV status (1.0% (-7.5%, 9.5%)) but the upper confidence limit is <10%, the maximal value allowed by the trial protocol to consider non-inferiority.
- j. The difference in proportions (-1.2% (-11.1%, 8.6%)) becomes smaller after adjustment for HIV status (-0.7% (-10.5%, 9.1%)).
- k. This outcome was a subset from among the "unfavourable outcomes" of the trial. "Unfavourable outcomes" were defined as not satisfying the favourable outcome criteria (see above) because of (i) start of 2 or more additional medicines (including a change of regimen); or (ii) treatment extended beyond the permitted duration; or (iii) death at any point up to 132 weeks post-randomization; or (iv) positive culture result at 132 weeks post-randomization or when last seen; or (v) not seen at 76 weeks or later
- I. In the mITT population the difference in mortality between shorter and longer regimen arms was also not statistically significant (9.5% vs 6.9% respectively; unadjusted RR 1.37 (0.66 to 2.86))
- m. Imprecision downgraded by two levels because the span of the confidence interval implies important uncertainty on potential risk of death or survival
- n. Time to death
- o. In patients without HIV the unadjusted RR for death was 0.610 (95%CI: 0.191-1.946). However, a subgroup analysis by HIV status was not pre-specified in the STREAM trial protocol.
- p. Unadjusted RR for lack of conversion, reversion or relapse in patients with baseline resistance to pyrazinamide was 5.95 (95% CL: 0.80, 44.15) and baseline resistance to ethionamide was 3.41 (95% CL: 0.20, 58.32)
- q. Moxifloxacin was given at 400mg/day in the longer regimen arm regardless of body weight; patients on the shorter regimen received either 400mg (<33kg body weight), 600mg (33-50kg) or 800mg (>50kg). A QTcF of 500ms+ was more frequent in shorter regimen than in longer regimen patients in both those weighing 33-50kg (9% vs 3%) and >50kg (10% vs 5%) (unadjusted RR 2.17 (95%CL: 0.92, 5.16)).
- r. The availability of electrocardiography in a typical centre treating MDR-TB patients may vary.
- s. Grade of severity was defined by the criteria of the Division of AIDS (DAIDS). Table for grading the severity of adult and pediatric adverse events. Version 1.0, Dec 2004; Clarification August 2009. http://rsc.tech-res.com/docs/default-source/safety/table for grading severity of adult pediatric adverse events.pdf. Hearing loss was not assessed using audiometry in all centres.
- t. The expert panel did not consider the variable use of audiometry to assess hearing loss by sites to be a reason to downgrade inconsistency for this outcome.
- u. The difference (-2.8% (-12.9%, 7.3%)) hardly changed after adjustment for HIV status (-2.8% (-12.9%, 7.2%))

Author(s): STREAM Stage 1 Trial investigators reported for the Guideline Development Group for the WHO treatment guidelines on MDR/RR-TB, 2018 update (6 July 2018) - FINAL RESULTS

Question: PICO 1. In MDR-TB patients, are treatment regimens lasting up to 12 months as likely to lead to the outcomes listed below when compared with those recommended in the WHO guidelines of 2011? (Cured/completed, culture conversion by 6 months, failure, relapse, survival (or death), adverse reactions, acquisition (amplification) of additional drug resistance, adherence to treatment interruption due to non-adherence)) [HIGH DOSE MFX/LFX RECODED FROM TRIAL ENDPOINTS TO WHO DEFINITIONS]

Setting: Assessment of final results from a phase III, multi-centre, international, parallel group, open-label, randomised controlled trial of a standardised MDR-TB treatment regimen lasting 9-11 months versus a longer regimen using a non-inferiority design. The trial enrolled patients between July 2012 and June 2015 in seven sites located in Ethiopia, Mongolia, South Africa and Viet Nam (intention to treat population = 424 [282 in study arm; 142 control arm]; mITT population = 369 [245 in study arm; 124 control arm]). Treatment allocation was not blinded to the participants, care givers or data managers. All local and reference laboratory assessments, including microbiological tests involved in the assignment of patient outcome, were conducted blind.

Bibliography: 1. ISRCTN - ISRCTN78372190: The evaluation of a standardised treatment regimen of anti-tuberculosis drugs for patients with multi-drug-resistant tuberculosis (MDR-TB). Available from: http://www.isrctn.com/ISRCTN78372190 2. Nunn AJ, Rusen ID, Van Deun A, Torrea G, Phillips PPJ, Chiang C-Y, et al. Evaluation of a standardized treatment regimen of anti-tuberculosis drugs for patients with multi-drug-resistant tuberculosis (STREAM): study protocol for a randomized controlled trial. Trials. 2014;15:353 3. Nunn A. Preliminary efficacy results of STREAM stage 1. In: STREAM trial (Evaluation of a Standardised Treatment Regimen of Anti-tuberculosis Drugs for Patients with Multidrug-resistant Tuberculosis): preliminary stage 1 results. Guadalajara, Mexico: UNION; 2017. Available from: http://www.professionalabstracts.com/theunion2017/iplanner/#/person/3892

			Certainty a	ssessment			N₂ of p	atients	Effec	l .	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a shorter treatment regimen (9-12 months)	longer regimens conforming to WHO guidelines	Relative (95% CI)	Absolute (95% CI)		
Fail/relapse v	vs Success: compa	ring Shorter regimen:	s using high-dose Mox	xifloxacin to Longer re	gimens using conven	tional dose fluoroquinolones						
1	randomised trials <sup>a</sup>	not serious	not serious	not serious	not serious	none	27/231 (11.7%)	8/117 (6.8%)	<b>RR 1.7</b> (0.8 to 3.6)	5 more per 100 (from 1 fewer to 11 more) <sup>b</sup>	⊕⊕⊕ HIGH	CRITICAL
Death vs Suc	ccess: comparing S	Shorter regimens usin	g high-dose Moxifloxa	cin to Longer regime	ns using conventional	dose fluoroquinolones						
1	randomised trials <sup>a</sup>	not serious	not serious	not serious	serious <sup>c</sup>	none	21/225 (9.3%)	8/117 (6.8%)	<b>RR 1.4</b> (0.6 to 3.0)	3 more per 100 (from 4 fewer to 8 more) <sup>b</sup>	⊕⊕⊕ MODERATE	CRITICAL
Lost vs Succe	ess, Fail/relapse: c	omparing Shorter reg	imens using high-dos	e Moxifloxacin to Lon	ger regimens using co	onventional dose fluoroquinolones	3					
1	randomised trials <sup>a</sup>	not serious	not serious	not serious	not serious	none	5/236 (2.1%)	6/123 (4.9%)	<b>RR 0.4</b> (0.1 to 1.4)	3 fewer per 100 (from 1 more to 0 fewer) <sup>b</sup>	⊕⊕⊕ ніGH	CRITICAL

CI: Confidence interval; RR: Risk ratio

- a. Because all data on high-dose fluoroquinolones in observational studies were only from two cohorts, and the majority were from a single cohort (Bangladesh), STREAM trial data were used for comparing a shorter regimen that utilized high-dose moxifloxacin (600mg to 800mg depending on body weight) to longer regimens utilizing usual-dose fluoroquinolones (moxifloxacin 400mg or levofloxacin 750-1000 mg), results of the comparison are shown here
- b. Unadjusted absolute risk difference shown, as there was balance in the distribution of confounders between the study arms of STREAM.
- c. Wide confidence-interval that includes no effect and potential harm.

Date: 16 July 2018

Question: PICO 1 In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) is a shorter treatment regimen (9-12 months) more or less likely to safely improve outcomes than longer regimens conforming to WHO guidelines?

Setting: Treatment of adults with MDR/RR-TB using shorter treatment regimens in observational studies in Bangladesh, Benin, Burkina Faso, Burundi, Cameroon, Côte d'Ivoire, Central African Republic, Democratic Republic of Congo, Eswatini, Kyrgyzstan, Niger, Rwanda, Tajikistan, South Africa and Uzbekistan

#### 1. overall findings

			Certainty a	ssessment			N₂ of p	atients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a shorter treatment regimen (9-12 months)	longer regimens conforming to WHO guidelines	Relative (95% CI)	Absolute (95% CI)		
Fail/relapse v	s Success: Shorte	r vs Longer regimens	5									
47	observational studies <sup>a,b,c</sup>	not serious <sup>d</sup>	not serious <sup>e</sup>	not serious	not serious	none	118/2282 (5.2%)	112/1926 (5.8%)	OR 2.00 (0.96 to 4.00)	2 more per 100 (from 0 more to 5 more)	⊕⊕⊖⊖ Low	CRITICAL
Death vs Suc	ccess: Shorter vs L	onger regimens										
46	observational studies <sup>a,b,c</sup>	not serious <sup>d</sup>	not serious <sup>e</sup>	serious <sup>f</sup>	not serious	none	201/2365 (8.5%)	265/2079 (12.7%)	<b>OR 1.20</b> (0.95 to 1.60)	2 more per 100 (from 1 fewer to 5 more)	⊕⊖⊖⊖ VERY LOW	CRITICAL
Lost vs Succ	ess, Fail/relapse: S	Shorter vs Longer reg	imens									
46	observational studies <sup>a,b,c</sup>	not serious <sup>d</sup>	not serious <sup>e</sup>	not serious	not serious	none	142/2424 (5.9%)	526/2452 (21.5%)	OR 0.2 (0.2 to 0.3)	15 fewer per 100 (from 17 fewer to 12 fewer)	⊕⊕⊖⊖ Low	CRITICAL
Fail/relapse v	s Success among	st patients with confir	med FQN-susceptible	TB: Shorter versus L	onger regimens that i	nclude Bdq, Lzd, or Delamanid						
35	observational studies	not serious <sup>d</sup>	not serious <sup>e</sup>	not serious	very serious <sup>g</sup>	strong association	86/1717 (5.0%)	14/337 (4.2%)	<b>OR 9.1</b> (0.6 to 135.4)	7 more per 100 (from 2 more to 12 more) h	VERY LOW	CRITICAL
Death vs Suc	cess amongst pati	ents with confirmed F	-QN-susceptible TB: S	Shorter versus Longer	regimens that includ	e Bdq, Lzd, or Delamanid				!		•
34	observational studies	not serious <sup>d</sup>	not serious <sup>e</sup>	serious <sup>f</sup>	serious <sup>g</sup>	none	146/1777 (8.2%)	34/357 (9.5%)	<b>OR 1.4</b> (0.7 to 2.5)	3 more per 100 (from 3 fewer to 8 more) h	VERY LOW	CRITICAL
Lost vs Succ	ess, Fail/relapse ar	mongst patients with	confirmed FQN-susce	ptible TB: Shorter ver	sus Longer regimens	that include Bdq, Lzd, or Delamai	nid					ļ.

			Certainty a	assessment			N₂ of p	atients	Effect	t	Certainty	Importance
N₂ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a shorter treatment regimen (9-12 months)	longer regimens conforming to WHO guidelines	Relative (95% CI)	Absolute (95% CI)		
37	observational studies	not serious <sup>d</sup>	not serious <sup>e</sup>	not serious	not serious	none	109/1826 (6.0%)	77/414 (18.6%)	<b>OR 0.4</b> (0.2 to 0.7)	10 fewer per 100 (from 16 fewer to 5 fewer) h	⊕⊕⊖⊖ Low	CRITICAL

CI: Confidence interval; OR: Odds ratio

- a. Fluoroquinolone dosing in Shorter regimens: usual-dose moxifloxacin (Mfx), n=1378 (6 cohorts); high-dose gatifloxacin (Gfx), n=897; usual-dose Gfx, n=246; high-dose levofloxacin (Lfx), n=206; usual-dose Lfx, n=1. Fluoroquinolone dosing in Longer regimens: usual-dose Mfx, n=2131; usual-dose Lfx, n=716; usual-dose Gfx, n=2. Gfx-containing Shorter regimens used in 3 cohorts: Bangladesh (n=829, 73% of Gfx patients), Cameroon (n=195, 17%), Niger(n=119, 10%). High-dose Gfx was used in 2 cohorts (N=778 in Bangladesh, N=119 in Niger). Comparison of Gfx-based Shorter regimens to Longer regimens (any later-generation fluoroquinolone): Failure/relapse versus Success: aOR 0.1 (0.0 to 3.5), aRD -4 (-6 to -2); for Death versus Success: aOR 0.8 (0.5 to 1.4), aRD -1 (-4 to 2); for Lost versus Success/failure/relapse: aOR 0.3 (0.2 to 0.4), aRD -14 (-18 to -10).
- b. Data on children & adolescents (age < 16 years old) were sparse (Shorter regimens, n=53 from 5 cohorts; Longer regimens, n=29 from 9 cohorts) and meta-analytic methods failed. Using simple pooling, unadjusted risk differences comparing Shorter to Longer regimens were: for failure/relapse vs success, -4 per 100 (95Cl: -10.2); for death vs success, +3 per 100 (95Cl: -3.10), lost vs success/failure/relapse, -14 per 100 (95Cl: -23,-4).
- c. No data available on pregnant patients.
- d. While residual confounding due to non-randomized design & between-study comparisons are possible given the observational nature of the data, the Guideline Development Group panel considered the methods utilized were the strongest approach to control for confounding amongst observational studies.
- e. The Guideline Development Group panel considered the issues of inconsistency and imprecision and the inability to potentially separate these two issues given the Propensity Score Matching. Given this, the panel recommended to not rate down inconsistency but to rate imprecision based on the usual criteria.
- f. Restricted to deaths during treatment for Shorter regimens, and deaths during first 12 months of treatment for longer regimens; whereas true outcome of interest for this comparison would be deaths over the same period of time following treatment initiation (post-treatment death for Shorter to achieve same duration of death amongst longer).
- g. Wide confidence-interval that includes no effect and potential harm.
- h. Adjusted risk difference estimated using fixed effects, propensity-score matched, IPD meta-analysis.

Date: 16 July 2018

Question: PICO 1 In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) is a shorter treatment regimen (9-12 months) more or less likely to safely improve outcomes than longer regimens conforming to WHO guidelines?

Setting: Treatment of adults with MDR/RR-TB using shorter treatment regimens in observational studies in Bangladesh, Benin, Burkina Faso, Burundi, Cameroon, Côte d'Ivoire, Central African Republic, Democratic Republic of Congo, Eswatini, Kyrgyzstan, Niger, Rwanda, Tajikistan, South Africa and Uzbekistan

#### 2. People living with HIV

			Certainty a	ssessment			Nº of p	atients	Effec	t	Certainty	Importance
N₂ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a shorter treatment regimen (9-12 months)	longer regimens conforming to WHO guidelines	Relative (95% CI)	Absolute (95% CI)		
Fail/relapse v	s Success in peop	ole living with HIV (PL	WH)									
15	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	not serious	serious <sup>c</sup>	none	24/295 (8.1%)	55/750 (7.3%)	<b>OR 2.1</b> (0.6 to 7.7)	3 more per 100 (from 2 fewer to 7 more)	⊕⊖⊖ VERY LOW	CRITICAL
Fail/relapse v	s Success in HIV-r	negative										•
47	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	not serious	not serious	none	94/1978 (4.8%)	56/1162 (4.8%)	OR 1.9 (0.9 to 4.0)	2 more per 100 (from 0 fewer to 4 more)	⊕⊕⊖⊖ Low	CRITICAL
Death vs Suc	ccess in PLWH											
14	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	serious <sup>d</sup>	serious <sup>c</sup>	none	72/343 (21.0%)	169/864 (19.6%)	OR 1.0 (0.6 to 1.6)	0 fewer per 100 (from 8 fewer to 8 more)	⊕⊖⊖⊖ VERY LOW	CRITICAL
Death vs Suc	ccess in HIV-negati	ive				•						
46	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	serious <sup>d</sup>	serious <sup>c</sup>	none	127/2011 (6.3%)	96/1202 (8.0%)	<b>OR 0.8</b> (0.4 to 1.4)	1 fewer per 100 (from 4 fewer to 1 more)	VERY LOW	CRITICAL
Lost vs Succ	ess, Fail/relapse in	PLWH										
15	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	not serious	not serious	very strong association	13/308 (4.2%)	237/987 (24.0%)	OR 0.1 (0.1 to 0.3)	20 fewer per 100 (from 28 fewer to 13 fewer)	<del>ПОВ</del>	CRITICAL
Lost vs Succ	ess, Fail/relapse in	HIV-negative	ı			l						

			Certainty a	assessment			N₂ of p	atients	Effect	ı	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a shorter treatment regimen (9-12 months)	longer regimens conforming to WHO guidelines	Relative (95% CI)	Absolute (95% CI)		
47	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	not serious	not serious	strong association	129/2107 (6.1%)	286/1448 (19.8%)	<b>OR 0.3</b> (0.2 to 0.4)	13 fewer per 100 (from 15 fewer to 10 fewer)	MODERATE	CRITICAL

CI: Confidence interval; OR: Odds ratio

- a. While residual confounding due to non-randomized design & between-study comparisons are possible given the observational nature of the data, the Guideline Development Group panel considered the methods utilized were the strongest approach to control for confounding amongst observational studies.
- b. The Guideline Development Group panel considered the issues of inconsistency and imprecision and the inability to potentially separate these two issues given the Propensity Score Matching. Given this, the panel recommended to not rate down inconsistency but to rate imprecision based on the usual criteria.
- c. Wide confidence-interval that includes no effect and potential harm.
- d. Restricted to deaths during treatment for Shorter regimens, and deaths during first 12 months of treatment for longer regimens; whereas true outcome of interest for this comparison would be deaths over the same period of time following treatment initiation (post-treatment death for Shorter to achieve same duration of death amongst longer).

Date: 16 July 2018

Question: PICO 1 In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) is a shorter treatment regimen (9-12 months) more or less likely to safely improve outcomes than longer regimens conforming to WHO guidelines?

Setting: Treatment of adults with MDR/RR-TB using shorter treatment regimens in observational studies in Bangladesh, Benin, Burkina Faso, Burundi, Cameroon, Côte d'Ivoire, Central African Republic, Democratic Republic of Congo, Eswatini, Kyrgyzstan, Niger, Rwanda, Tajikistan, South Africa and Uzbekistan

#### 3. Extent of disease

			Certainty a	ssessment			Nº of p	patients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a shorter treatment regimen (9-12 months)	longer regimens conforming to WHO guidelines	Relative (95% CI)	Absolute (95% CI)		
Fail/relapse v	s Success Extensi	ive	-									
45	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	serious <sup>c</sup>	serious <sup>d</sup>	none	91/1969 (4.6%)	83/1320 (6.3%)	<b>OR 1.2</b> (0.6 to 2.6)	1 more per 100 (from 2 fewer to 3 more) <sup>e</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Fail/relapse v	s Success Not ext	ensive										
34	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	not serious	not serious	none	20/259 (7.7%)	28/602 (4.7%)	OR 2.9 (0.8 to 10.0)	5 more per 100 (from 0 fewer to 9 more) <sup>e</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs Suc	ccess Extensive											
44	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	serious <sup>c</sup>	not serious	none	165/2043 (8.1%)	185/1422 (13.0%)	OR 1.0 (0.8 to 1.3)	0 fewer per 100 (from 3 fewer to 3 more) <sup>e</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Death vs Suc	ccess Not extensive	e										
36	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	serious °	serious <sup>e</sup>	none	27/266 (10.2%)	79/653 (12.1%)	<b>OR 1.6</b> (0.5 to 5.6)	2 more per 100 (from 4 fewer to 7 more) <sup>e</sup>	⊕⊖⊖⊖ VERY LOW	CRITICAL
Lost vs Succe	ess, Fail/relapse E	xtensive										
46	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	not serious	not serious	strong association	122/2091 (5.8%)	368/1688 (21.8%)	<b>OR 0.3</b> (0.2 to 0.4)	15 fewer per 100 (from 18 fewer to 12 fewer) °	MODERATE	CRITICAL
Lost vs Succe	ess, Fail/relapse N	ot extensive	l			L						L

			Certainty a	assessment			Nº of p	atients	Effec	1	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a shorter treatment regimen (9-12 months)	longer regimens conforming to WHO guidelines	Relative (95% CI)	Absolute (95% CI)		
35	observational studies	not serious <sup>a</sup>	not serious <sup>b</sup>	not serious	not serious	strong association	16/275 (5.8%)	157/759 (20.7%)	OR 0.3 (0.1 to 0.5)	13 fewer per 100 (from 20 fewer to 6 fewer) °	⊕⊕⊕ MODERATE	CRITICAL

CI: Confidence interval; OR: Odds ratio

- a. While residual confounding due to non-randomized design & between-study comparisons are possible given the observational nature of the data, the Guideline Development Group panel considered the methods utilized were the strongest approach to control for confounding amongst observational studies.
- b. The Guideline Development Group panel considered the issues of inconsistency and imprecision and the inability to potentially separate these two issues given the Propensity Score Matching. Given this, the panel recommended to not rate down inconsistency but to rate imprecision based on the usual criteria.
- c. Restricted to deaths during treatment for Shorter regimens, and deaths during first 12 months of treatment for longer regimens; whereas true outcome of interest for this comparison would be deaths over the same period of time following treatment initiation (post-treatment death for Shorter to achieve same duration of death amongst longer).
- d. Wide confidence-interval that includes no effect and potential harm.
- e. Adjusted risk difference estimated using fixed effects, propensity-score matched, IPD meta-analysis.

Date: 16 July 2018

Question: PICO 1 In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) is a shorter treatment regimen (9-12 months) more or less likely to safely improve outcomes than longer regimens conforming to WHO guidelines?

Setting: Treatment of adults with MDR/RR-TB using shorter treatment regimens in observational studies in Bangladesh, Benin, Burkina Faso, Burundi, Cameroon, Côte d'Ivoire, Central African Republic, Democratic Republic of Congo, Eswatini, Kyrgyzstan, Niger, Rwanda, Tajikistan, South Africa and Uzbekistan

#### 4. Pyrazinamide and ethionamide resistance

			Certainty a	assessment			N₂ of p	atients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a shorter treatment regimen (9-12 months)	longer regimens conforming to WHO guidelines	Relative (95% CI)	Absolute (95% CI)		
Fail/relapse v	s Success PZA-Re	esistant, FQ-S										
31	observational studies	not serious <sup>a,b</sup>	not serious <sup>c</sup>	not serious	not serious	strong association	36/270 (13.3%)	11/349 (3.2%)	OR 10.7 (1.8 to 64.5)	12 more per 100 (from 7 more to 16 more) <sup>d</sup>	⊕⊕⊕⊖ MODERATE	CRITICAL
Fail/relapse v	s Success PZA-Su	usceptible, FQ-S				•						
28	observational studies	not serious <sup>a</sup>	not serious <sup>c</sup>	not serious	serious <sup>e</sup>	none	12/248 (4.8%)	13/428 (3.0%)	<b>OR 1.3</b> (0.3 to 6.7)	1 more per 100 (from 4 fewer to 5 more) <sup>d</sup>	₩ VERY LOW	CRITICAL
Death vs Suc	ccess PZA-Resista	nt, FQ-S	•			<u>'</u>				'		
32	observational studies	not serious <sup>a</sup>	not serious <sup>c</sup>	serious <sup>1</sup>	not serious	none	16/250 (6.4%)	33/371 (8.9%)	OR 0.3 (0.1 to 1.4)	5 fewer per 100 (from 11 fewer to 2 more) <sup>d</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Death vs Suc	ccess PZA-Suscept	tible, FQ-S										
27	observational studies	not serious <sup>a</sup>	not serious <sup>c</sup>	serious <sup>1</sup>	very serious <sup>e</sup>	none	19/255 (7.5%)	19/434 (4.4%)	<b>OR 1.4</b> (0.4 to 5.5)	1 more per 100 (from 5 fewer to 7 more) <sup>d</sup>	⊕⊖⊖⊖ VERY LOW	CRITICAL
Lost vs Succ	ess, Fail/relapse, F	ZA-Resistant, FQ-S	ļ.	!	<u>l</u>	<b>!</b>						<b>!</b>
33	observational studies	not serious <sup>a</sup>	not serious <sup>c</sup>	not serious	not serious	none	13/283 (4.6%)	54/403 (13.4%)	<b>OR 0.2</b> (0.0 to 1.4)	10 fewer per 100 (from 16 fewer to 4 fewer) d	⊕⊕⊖⊖ Low	CRITICAL
Lost vs Succ	L ess, Fail/relapse, F	ZA-Susceptible, FQ-	S									L

			Certainty a	assessment			Nº of p	atients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a shorter treatment regimen (9-12 months)	longer regimens conforming to WHO guidelines	Relative (95% CI)	Absolute (95% CI)		
30	observational studies	not serious <sup>a</sup>	not serious <sup>c</sup>	not serious	not serious	strong association	17/265 (6.4%)	103/531 (19.4%)	OR 0.3 (0.1 to 0.5)	15 fewer per 100 (from 22 fewer to 7 fewer) <sup>d</sup>	MODERATE	CRITICAL
Fail/relapse v	vs Success ETO/P	TO-Resistance, FQ-S	5	<u> </u>		l				1		
31	observational studies	not serious <sup>a,b</sup>	not serious °	not serious	not serious	none	23/249 (9.2%)	4/149 (2.7%)	OR 3.9 (1.0 to 15.1)	7 more per 100 (from 1 fewer to 16 more) <sup>d</sup>	⊕⊕⊖⊖ Low	CRITICAL
Fail/relapse \	vs Success ETO/P	TO-Susceptible, FQ-S	5	Į.			<b>'</b>					
34	observational studies	not serious <sup>a</sup>	not serious <sup>c</sup>	not serious	serious <sup>e</sup>	none	7/660 (1.1%)	26/613 (4.2%)	<b>OR 0.1</b> (0.0 to 1.5)	3 fewer per 100 (from 5 fewer to 1 fewer) d	VERY LOW	CRITICAL
Death vs Suc	L ccess ETO/PTO-Re	esistance, FQ-S				L			<u> </u>			
30	observational studies	not serious <sup>a</sup>	not serious <sup>c</sup>	serious <sup>1</sup>	serious <sup>e</sup>	none	18/244 (7.4%)	9/154 (5.8%)	<b>OR 1.5</b> (0.3 to 7.4)	2 more per 100 (from 5 fewer to 8 more) d	WERY LOW	CRITICAL
Death vs Suc	L ccess ETO/PTO-Su	Lusceptible, FQ-S										
34	observational studies	not serious <sup>a</sup>	not serious <sup>c</sup>	serious <sup>1</sup>	very serious <sup>e</sup>	none	61/714 (8.5%)	36/623 (5.8%)	OR 2.1 (0.8 to 5.8)	2 more per 100 (from 1 fewer to 6 more) <sup>d</sup>	₩ CONTRACTOR	CRITICAL
Lost vs Succ	ess, Fail/relapse E	TO/PTO-Resistance,	FQ-S									
32	observational studies	not serious <sup>a</sup>	not serious <sup>c</sup>	not serious	not serious	very strong association	13/262 (5.0%)	42/191 (22.0%)	<b>OR 0.1</b> (0.0 to 0.4)	19 fewer per 100 (from 26 fewer to 12 fewer) <sup>d</sup>	HIGH	CRITICAL
Lost vs Succ	ess, Fail/relapse E	I TO/PTO-Susceptible	, FQ-S									
32	observational studies	not serious <sup>a</sup>	not serious <sup>c</sup>	not serious	not serious	strong association	53/713 (7.4%)	122/735 (16.6%)	<b>OR 0.4</b> (0.3 to 0.7)	7 fewer per 100 (from 11 fewer to 4 fewer) <sup>d</sup>	⊕⊕⊕⊖ MODERATE	CRITICAL

CI: Confidence interval; OR: Odds ratio

- a. While residual confounding due to non-randomized design & between-study comparisons are possible given the observational nature of the data, the Guideline Development Group panel considered the methods utilized were the strongest approach to control for confounding amongst observational studies.
- b. Adjusted odds ratio estimated from fixed-effects meta-analysis due to non-convergence of mixed-effects models.
- c. The Guideline Development Group panel considered the issues of inconsistency and imprecision and the inability to potentially separate these two issues given the Propensity Score Matching. Given this, the panel recommended to not rate down inconsistency but to rate imprecision based on the usual criteria.
- d. Adjusted risk difference estimated using fixed effects, propensity-score matched, IPD meta-analysis.
- e. Wide confidence-interval that includes no effect and potential harm.
- f. Restricted to deaths during treatment for Shorter regimens, and deaths during first 12 months of treatment for longer regimens; whereas true outcome of interest for this comparison would be deaths over the same period of time following treatment initiation (post-treatment death for Shorter to achieve same duration of death amongst longer).

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

#### 1. Levofloxacin / moxifloxacin

			Certainty a	ssessment			N₂ of p	patients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	levofloxacin or moxifloxacin	no fluoroquinolones or a FQ of different generation (Ofx/Mfx/Lfx/Gfx)	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	lure/relapse vs. tre	atment success - Lev	ofloxacin or Moxifloxa	acin (S) vs NO FQ (all	)							
53	observational studies	not serious	not serious	not serious	not serious	strong association	226/3143 (7.2%)	67/275 (24.4%)	<b>aOR 0.3</b> (0.1 to 0.5)	16 fewer per 100 (from 23 fewer to 9 fewer) <sup>a</sup>	MODERATE	CRITICAL
Death vs. trea	atment success - L	evofloxacin or Moxiflo	oxacin (S) vs NO FQ (	all)		•						
53	observational studies	not serious	not serious	not serious	not serious	strong association	634/3551 (17.9%)	208/416 (50.0%)	<b>aOR 0.2</b> (0.1 to 0.3)	35 fewer per 100 (from 42 fewer to 29 fewer) <sup>a</sup>	MODERATE	CRITICAL
Treatment fai	lure/relapse vs. tre	atment success - Lev	ofloxacin or Moxifloxa	acin (S) vs Ofloxacin (	S)							
53	observational studies	not serious	not serious	not serious	not serious	none	226/3143 (7.2%)	199/1273 (15.6%)	<b>aOR 0.4</b> (0.3 to 0.6)	8 fewer per 100 (from 11 fewer to 5 fewer) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - L	evofloxacin or Moxiflo	oxacin (S) vs Ofloxaci	n (S)	<u>I</u>							
53	observational studies	not serious	not serious	not serious	not serious	none	634/3551 (17.9%)	288/1362 (21.1%)	<b>aOR 0.8</b> (0.6 to 1.0)	3 fewer per 100 (from 7 fewer to 0 fewer) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Sensitivity An	alysis - Treatment	failure/relapse vs. tre	eatment success - Lev	ofloxacin or Moxifloxa	cin (S or U) vs Ofloxa	acin (R)						
53	observational studies	not serious	not serious	not serious	not serious	none	117/703 (16.6%)	10/54 (18.5%)	<b>aOR 0.5</b> (0.0 to 28.0)	3 fewer per 100 (from 22 fewer to 17 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL

			Certainty a	ssessment			N₂ of p	patients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	levofloxacin or moxifloxacin	no fluoroquinolones or a FQ of different generation (Ofx/Mfx/Lfx/Gfx)	Relative (95% CI)	Absolute (95% CI)		
Sensitivity An	alysis - Death vs. 1	reatment success - L	evofloxacin or Moxiflo	xacin (S or U) vs Oflo	oxacin (R)					•		
53	observational studies	not serious	not serious	not serious	not serious	none	195/781 (25.0%)	36/80 (45.0%)	<b>aOR 0.5</b> (0.2 to 1.2)	13 fewer per 100 (from 31 fewer to 4 more) <sup>a</sup>	⊕⊕⊖ Low	CRITICAL
Sensitivity An	alysis - in XDR, Tr	eatment failure/relaps	se vs. treatment succe	ess - Levofloxacin or N	Moxifloxacin (all) vs N	O FQ (all)						
53	observational studies	not serious	not serious	not serious	serious <sup>b</sup>	none	128/640 (20.0%)	68/221 (30.8%)	<b>aOR 0.7</b> (0.4 to 1.3)	7 fewer per 100 (from 18 fewer to 5 more) <sup>a</sup>	VERY LOW	CRITICAL
Sensitivity An	alysis - in XDR, de	eath vs. treatment suc	ccess - Levofloxacin o	r Moxifloxacin (all) vs	NO FQ (all)					!		
53	observational studies	not serious	not serious	not serious	serious <sup>b</sup>	none	229/741 (30.9%)	261/414 (63.0%)	<b>aOR 0.5</b> (0.3 to 0.7)	15 fewer per 100 (from 23 fewer to 7 fewer) <sup>a</sup>	VERY LOW	CRITICAL
Adverse ever	nts (levofloxacin dr	ug stopped permanei	ntly)			ļ.				!		l
17	observational studies	not serious	not serious °	not serious	not serious	none	19/883 (2.2%)		not estimable	1 more per 100 (from 0 more to 6 more) <sup>d</sup>	⊕⊕⊖⊖ Low	IMPORTANT
Adverse ever	nts (moxifloxacin di	rug stopped permane	ently)									
22	observational studies	not serious	not serious	not serious	not serious	none	27/826 (3.3%)		not estimable	3 more per 100 (from 2 more to 5 more) <sup>d</sup>	⊕⊕⊖⊖ Low	

CI: Confidence interval

- a. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- b. Small numbers
- c. Simulated I squared from one-step IPD exceeds 50%
- d. Pooled incidence of adverse events of random effect in meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

#### 2. Bedaquiline (adults)

			Certainty a	assessment			N₂ of p	patients	Effec	et	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	bedaquiline	no bedaquiline	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	ilure/relapse vs tre	atment success - Stra	ain susceptible to Bed	aquiline						'		
53	observational studies	not serious	not serious	not serious	not serious	strong association	133/1391 (9.6%)	419/3237 (12.9%)	<b>aOR 0.3</b> (0.2 to 0.4)	17 fewer per 100 (from 23 fewer to 11 fewer) <sup>a</sup>	MODERATE	CRITICAL
Death vs trea	tment success - S	train susceptible to B	edaquiline	'		<u> </u>	<u> </u>			, ,		
53	observational studies	not serious	not serious	not serious	not serious	strong association	222/1480 (15.0%)	831/3649 (22.8%)	<b>aOR 0.2</b> (0.2 to 0.3)	30 fewer per 100 (from 34 fewer to 26 fewer) <sup>a</sup>	MODERATE	CRITICAL
RCT, TMC20	7-C208 - Treatmer	nt failure/relapse vs tr	eatment success	L	L		<b>'</b>					
1	randomised trials	not serious	not serious	not serious	serious <sup>b</sup>	none	5/43 (11.6%)	20/41 (48.8%)	<b>aOR 0.1</b> (0.0 to 0.4)	37 fewer per 100 (from 55 fewer to 19 fewer) <sup>a</sup>	MODERATE	CRITICAL
RCT, TMC20	7-C208 - Death vs	treatment success				L	L			<u> </u>		L
1	randomised trials	not serious	not serious	not serious	serious <sup>b</sup>	none	8/46 (17.4%)	2/23 (8.7%)	<b>aOR 5.0</b> (0.6 to 42.0)	15 more per 100 (from 0 fewer to 29 more) <sup>a</sup>	⊕⊕⊕⊖ MODERATE	CRITICAL
Sensitivity An	nalysis - Case-cont	rol study using South	Africa data only - Trea	atment failure/relapse	vs treatment succes	s - Restricted to strains susceptible	e to Bedaquiline					•
53	observational studies	not serious	not serious	not serious	not serious	strong association	69/869 (7.9%)	276/1188 (23.2%)	<b>aOR 0.3</b> (0.2 to 0.4)	15 fewer per 100 (from 18 fewer to 12 fewer) <sup>a</sup>	⊕⊕⊕ MODERATE	CRITICAL
Sensitivity An	I nalysis - Case-cont	I rol study using South	Africa data only - Dea	I ath vs treatment succe	ess - Restricted to str	I ains susceptible to Bedaquiline						

			Certainty a	assessment			Nº of p	patients	Effec	t	Certainty	Importance
Nº of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	bedaquiline	no bedaquiline	Relative (95% CI)	Absolute (95% CI)		
53	observational studies	not serious	not serious	not serious	not serious	strong association	176/976 (18.0%)	694/1606 (43.2%)	<b>aOR 0.3</b> (0.2 to 0.4)	25 fewer per 100 (from 29 fewer to 22 fewer) <sup>a</sup>	MODERATE	CRITICAL
Sensitivity Ar	nalysis - Patients w	ith XDR (No restriction	n on number of effect	ive drugs) - Treatmen	t failure/relapse vs tre	eatment success						
53	observational studies	not serious	not serious	not serious	serious <sup>b</sup>	strong association	59/512 (11.5%)	191/539 (35.4%)	<b>aOR 0.2</b> (0.1 to 0.3)	26 fewer per 100 (from 36 fewer to 16 fewer) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Sensitivity Ar	nalysis - Patients w	ith XDR (No restriction	n on number of effect	ive drugs) - Death vs	treatment success							
53	observational studies	not serious	not serious	not serious	serious <sup>b</sup>	strong association	98/551 (17.8%)	469/817 (57.4%)	<b>aOR 0.1</b> (0.1 to 0.2)	45 fewer per 100 (from 52 fewer to 39 fewer) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Sensitivity Ar	nalysis - Excluding	patients who received	d other new drugs (Lz	d, CFZ or Carbapene	ms), no restriction on	number of effective drugs - Treatn	nent failure/relapse vs tre	eatment success				
53	observational studies	not serious	not serious	not serious	not serious	strong association	40/415 (9.6%)	621/4219 (14.7%)	<b>aOR 0.3</b> (0.2 to 0.5)	16 fewer per 100 (from 24 fewer to 8 fewer) <sup>a</sup>	MODERATE	CRITICAL
Sensitivity Ar	nalysis - Excluding	patients who received	d other new drugs (Lz	d, CFZ or Carbapene	ms), no restriction on	number of effective drugs - Death	vs treatment success					<u> </u>
53	observational studies	not serious	not serious	not serious	not serious	strong association	67/442 (15.2%)	1127/4725 (23.9%)	<b>aOR 0.3</b> (0.2 to 0.4)	24 fewer per 100 (from 31 fewer to 17 fewer) <sup>a</sup>	MODERATE	CRITICAL
Adverse ever	nts (drug stopped p	permanently)										
12	observational studies	not serious	not serious	not serious	not serious	none	8/447 (1.8%)		not estimable	2 more per 100 (from 1 more to 4 more) °	⊕⊕⊖⊖ <sub>Low</sub>	IMPORTANT

CI: Confidence interval

- $a.\ Risk\ differences\ and\ confidence\ interval\ from\ propensity\ score\ matched\ regression\ meta-analysis\ of\ individual\ patient\ records$
- b. Small numbers



Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

#### 3. Linezolid

			Certainty a	assessment			N₂ofp	atients	Effec	t	Certainty	Importance
N₂ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	linezolid	no linezolid	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	lure/relapse vs. tre	eatment success - Str	ain susceptible to Line	ezolid						'		
53	observational studies	not serious	not serious <sup>a</sup>	not serious	not serious	strong association	118/1216 (9.7%)	409/2725 (15.0%)	<b>aOR 0.3</b> (0.2 to 0.5)	16 fewer per 100 (from 23 fewer to 8 fewer) <sup>b</sup>	⊕⊕⊕○ MODERATE	CRITICAL
Death vs. trea	atment success - S	Strain susceptible to L	inezolid									
53	observational studies	not serious	not serious <sup>a</sup>	not serious	not serious	strong association	188/1286 (14.6%)	790/3106 (25.4%)	<b>aOR 0.3</b> (0.2 to 0.3)	25 fewer per 100 (from 31 fewer to 19 fewer) <sup>b</sup>	MODERATE	CRITICAL
Sensitivity An	nalysis - Patients w	ith XDR (no restriction	n on number of effecti	ve drugs) - Treatment	failure/relapse vs. tre	eatment success				<u> </u>		<b>'</b>
53	observational studies	not serious	not serious	not serious	not serious	strong association	63/599 (10.5%)	187/451 (41.5%)	<b>aOR 0.2</b> (0.1 to 0.3)	31 fewer per 100 (from 38 fewer to 24 fewer) <sup>b</sup>	⊕⊕⊕ MODERATE	CRITICAL
Sensitivity An	lalysis - Patients w	I ith XDR (no restrictio	n on number of effecti	ve drugs) - Death vs.	treatment success							
53	observational studies	not serious	not serious	not serious	not serious	strong association	98/634 (15.5%)	466/730 (63.8%)	<b>aOR 0.2</b> (0.1 to 0.2)	35 fewer per 100 (from 41 fewer to 28 fewer) <sup>b</sup>	⊕⊕⊕ MODERATE	CRITICAL
Sensitivity An	nalysis - Excluding	patients who received	d other new drugs (BD	I OQ, CFZ, Carbapenen	ns), no restriction on I	number of effective drugs - Treatm	nent failure/relapse vs. tre	eatment success				l
53	observational studies	not serious	not serious	not serious	not serious	strong association	29/315 (9.2%)	570/3427 (16.6%)	<b>aOR 0.2</b> (0.1 to 0.4)	20 fewer per 100 (from 29 fewer to 12 fewer) <sup>b</sup>	⊕⊕⊕ MODERATE	CRITICAL
Sensitivity An	nalysis - Excluding	patients who received	d other new drugs (BD	Q, CFZ, Carbapenen	ns), no restriction on I	number of effective drugs - Death	vs. treatment success					

			Certainty a	assessment			Nº of p	atients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	linezolid	no linezolid	Relative (95% CI)	Absolute (95% CI)		
53	observational studies	not serious	not serious	not serious	not serious	strong association	31/317 (9.8%)	1055/3912 (27.0%)	<b>aOR 0.5</b> (0.3 to 0.9)	7 fewer per 100 (from 13 fewer to 1 fewer) <sup>b</sup>	MODERATE	CRITICAL
Sensitivity Ar	nalysis - LZD at dos	se of 600mg/day only	- Treatment failure/re	lapse vs. treatment su	uccess - Restricted to	strain susceptible to Linezolid						<u> </u>
53	observational studies	not serious	not serious	not serious	not serious	strong association	95/976 (9.7%)	409/2725 (15.0%)	<b>aOR 0.3</b> (0.2 to 0.5)	16 fewer per 100 (from 23 fewer to 8 fewer) <sup>b</sup>	⊕⊕⊕○ MODERATE	CRITICAL
Sensitivity Ar	nalysis - LZD at dos	se of 600mg/day only	- Death vs. treatment	success - Restricted	to strain susceptible t	to Linezolid						
53	observational studies	not serious	not serious	not serious	not serious	strong association	143/1024 (14.0%)	790/3106 (25.4%)	<b>aOR 0.2</b> (0.2 to 0.3)	26 fewer per 100 (from 32 fewer to 19 fewer) <sup>b</sup>	MODERATE	CRITICAL
Sensitivity Ar	nalysis - LZD given	for at least 5.5 month	ns - Treatment failure/	relapse vs. treatment	success - Restricted	to strain susceptible to Linezolid						
53	observational studies	not serious	not serious	not serious	not serious	strong association	45/303 (14.9%)	147/1116 (13.2%)	<b>aOR 0.60</b> (0.40 to 0.99)	7 fewer per 100 (from 15 fewer to 0 fewer) b	MODERATE	CRITICAL
Sensitivity Ar	nalysis - LZD given	for at least 5.5 month	ns - Death vs. treatme	nt success - Restricte	ed to strain susceptible	e to Linezolid						<b>!</b>
53	observational studies	not serious	not serious	not serious	not serious	strong association	16/274 (5.8%)	339/1308 (25.9%)	<b>aOR 0.3</b> (0.1 to 0.5)	12 fewer per 100 (from 18 fewer to 6 fewer) b	⊕⊕⊕○ MODERATE	CRITICAL
Adverse ever	nts (drug stopped p	permanently)				1						
31	observational studies	not serious	not serious <sup>a</sup>	not serious	not serious	strong association	130/724 (18.0%)	0/1000 (0.0%)	not estimable	14 more per 100 (from 0 fewer to 0 fewer) °	⊕⊕⊕○ MODERATE	IMPORTANT

CI: Confidence interval; OR: Odds ratio

- a. Simulated I squared from one-step IPD exceeds 50%
- b. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records



Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

#### 4. Clofazimine

			Certainty a	assessment			N₂ofp	atients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	clofazimine	no clofazimine	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	lure/relapse vs. tre	eatment success - Str	ain susceptible to Clo	fazimine			•			'		
53	observational studies	not serious	serious <sup>a</sup>	not serious	not serious	none	107/991 (10.8%)	421/3035 (13.9%)	<b>aOR 0.3</b> (0.2 to 0.5)	18 fewer per 100 (from 28 fewer to 8 fewer) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Death vs. tre	atment success - S	Strain susceptible to C	Clofazimine	Į.			<b>'</b>					
53	observational studies	not serious	serious <sup>a</sup>	not serious	not serious	none	212/1096 (19.3%)	731/3345 (21.9%)	<b>aOR 0.4</b> (0.3 to 0.6)	16 fewer per 100 (from 24 fewer to 8 fewer) b	₩ CONTRACTOR VERY LOW	CRITICAL
RCT, Tang et	al CID2015 - Trea	tment failure/relapse	vs. treatment success	- Strain susceptible t	to Clofazimine		<b>'</b>		ļ			
1	randomised trials	serious <sup>c</sup>	not serious	not serious	very serious <sup>d</sup>	none	6/45 (13.3%)	15/43 (34.9%)	<b>aOR 0.3</b> (0.1 to 0.8)	22 fewer per 100 (from 39 fewer to 4 fewer) <sup>b</sup>	VERY LOW	CRITICAL
RCT, Tang et	al CID2015 - Deat	th vs. treatment succe	ess - Strain susceptibl	e to Clofazimine		L	L		<u> </u>	<u> </u>		L
1	randomised trials	serious <sup>c</sup>	not serious	not serious	very serious <sup>d</sup>	none	4/43 (9.3%)	4/32 (12.5%)	<b>aOR 0.7</b> (0.2 to 3.1)	3 fewer per 100 (from 18 fewer to 11 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Sensitivity Ar	nalysis - Patients w	I ith XDR (no restrictio	n on number of effecti	ve drugs) - Treatment	t failure/relapse vs. tr	eatment success	L					
53	observational studies	not serious	not serious	not serious	serious <sup>d</sup>	none	63/417 (15.1%)	187/630 (29.7%)	<b>aOR 0.60</b> (0.30 to 1.01)	8 fewer per 100 (from 16 fewer to 1 more) <sup>b</sup>	₩ VERY LOW	CRITICAL
Sensitivity Ar	l nalysis - Patients w	I ith XDR (no restrictio	n on number of effecti	Ve drugs) - Death vs.	treatment success		l					

			Certainty a	assessment			№ of p	atients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	clofazimine	no clofazimine	Relative (95% CI)	Absolute (95% CI)		
53	observational studies	not serious	not serious	not serious	serious <sup>d</sup>	strong association	103/457 (22.5%)	464/907 (51.2%)	<b>aOR 0.4</b> (0.3 to 0.6)	17 fewer per 100 (from 24 fewer to 10 fewer) <sup>b</sup>	⊕⊕⊖⊖ Low	CRITICAL
Sensitivity Ar	nalysis - Excluding	patients who received	d other new drugs (BD	OQ, Lzd, Carbapenem	s), no restriction on n	umber of effective drugs - Treatme	ent failure/relapse vs. trea	atment success				
53	observational studies	not serious	not serious	not serious	serious <sup>e</sup>	none	47/227 (20.7%)	611/3958 (15.4%)	<b>aOR 1.7</b> (0.3 to 9.1)	5 more per 100 (from 6 fewer to 16 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Sensitivity Ar	nalysis - Excluding	patients who received	d other new drugs (BD	OQ, Lzd, Carbapenem	s), no restriction on n	umber of effective drugs - Death v	s. treatment success					
53	observational studies	not serious	not serious	not serious	serious <sup>e</sup>	none	94/274 (34.3%)	1009/4356 (23.2%)	<b>aOR 1.70</b> (0.96 to 3.10)	9 more per 100 (from 1 fewer to 20 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Adverse ever	nts (drug permaner	ntly stopped)										
11	observational studies	not serious	serious <sup>†</sup>	not serious	not serious	none	12/1689 (0.7%)		not estimable	2 more per 100 (from 1 more to 6 more) <sup>g</sup>	⊕⊖⊖ VERY LOW	IMPORTANT

CI: Confidence interval

- a. The beneficial effect of clofazimine was not observed in the sensitivity analysis excluding patients who received bedaquiline, carbapenems or linezolid.
- b. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- c. Randomization method unclear
- d. Small numbers
- e. A wide confidence interval was noted.
- f. Simulated I squared from one-step IPD exceeds 50%
- g. Pooled incidence of adverse events of random effect in meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

#### 5. Cycloserine / terizidone

			Certainty a	ssessment			N₂ of p	atients	Effec	ı	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	cycloserine or terizidone	no cycloserine or terizidone	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	ilure/relapse vs. tre	atment success - Stra	ain susceptible to Cyc	loserine or Terizidone								
53	observational studies	not serious	not serious	not serious	not serious	none	567/5483 (10.3%)	135/764 (17.7%)	<b>aOR 0.6</b> (0.4 to 0.9)	6 fewer per 100 (from 11 fewer to 2 fewer) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - S	Strain susceptible to C	cycloserine or Terizido	ne								1
53	observational studies	not serious	not serious	not serious	not serious	none	1244/6160 (20.2%)	274/903 (30.3%)	<b>aOR 0.6</b> (0.5 to 0.8)	9 fewer per 100 (from 13 fewer to 4 fewer) <sup>a</sup>	⊕⊕⊖ Low	CRITICAL
Sensitivity Ar	nalysis - Excluding	patients who received	i LZD, BDQ, CFZ or (	Carbapenems, no rest	riction on number of e	effective drugs - Treatment failure/	relapse vs treatment suc	cess		'		<u> </u>
53	observational studies	not serious	not serious	not serious	not serious	none	596/4904 (12.2%)	153/856 (17.9%)	<b>aOR 0.70</b> (0.50 to 0.99)	4 fewer per 100 (from 8 fewer to 0 fewer) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Sensitivity Ar	nalysis - Excluding	patients who received	I LZD, BDQ, CFZ or (	Carbapenems, no rest	riction on number of e	effective drugs - Death vs treatmen	nt success					1
53	observational studies	not serious	not serious	not serious	not serious	none	1184/5492 (21.6%)	405/1108 (36.6%)	<b>aOR 0.7</b> (0.5 to 0.9)	7 fewer per 100 (from 12 fewer to 3 fewer) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Adverse ever	nts (drugs stopped	permanently)										
35	observational studies	not serious	serious <sup>b</sup>	not serious	not serious	none	331/7374 (4.5%)		not estimable	6 more per 100 (from 4 more to 8 more) °	⊕⊖⊖ VERY LOW	IMPORTANT

CI: Confidence interval

- a. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- b. Simulated I squared from one-step IPD exceeds 50%
- c. Pooled incidence of adverse events of random effect in meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

#### 6. Ethambutol

		Certainty a	ssessment			N₂ of p	atients	Effec	t	Certainty	Importance
Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	ethambutol	no ethambutol	Relative (95% CI)	Absolute (95% CI)		
ilure/relapse vs trea	atment success - Stra	in susceptible to Etha	mbutol								
observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	123/1163 (10.6%)	50/411 (12.2%)	<b>aOR 0.40</b> (0.10 to 0.99)	7 fewer per 100 (from 11 fewer to 2 fewer) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
atment success - Si	train susceptible to E	thambutol									
observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	205/1245 (16.5%)	55/416 (13.2%)	<b>aOR 0.5</b> (0.1 to 1.7)	5 fewer per 100 (from 11 fewer to 1 more) b	⊕⊖⊖ VERY LOW	CRITICAL
ilure/relapse vs trea	atment success - Stra	in resistant to Ethamb	outol								
observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	64/424 (15.1%)	228/1974 (11.6%)	<b>aOR 1.3</b> (0.7 to 2.3)	3 more per 100 (from 3 fewer to 9 more) <sup>b</sup>	₩ VERY LOW	CRITICAL
atment success - Si	train resistant to Etha	mbutol			L						
observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	122/482 (25.3%)	347/2093 (16.6%)	<b>aOR 1.6</b> (1.1 to 2.4)	8 more per 100 (from 2 more to 14 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
nts (drug permaner	ntly stopped)										
observational studies	not serious	serious °	not serious	serious <sup>a</sup>	none	121/5954 (2.0%)		not estimable	2 more per 100 (from 1 more to 3 more) <sup>d</sup>	⊕⊖⊖⊖ VERY LOW	IMPORTANT
	observational studies  observational studies  thment success - So observational studies  lure/relapse vs treations observational studies  observational studies  observational studies  observational studies  observational studies	observational studies not serious  observational not serious  observational not serious	Study design   Risk of bias   Inconsistency	continue   continue	Study design   Risk of bias   Inconsistency   Indirectness   Imprecision	Study design   Risk of bias   Inconsistency   Indirectness   Imprecision   Other considerations	Study design   Risk of bias   Inconsistency   Indirectness   Imprecision   Other considerations   ethambutol	Study design Risk of bias Inconsistency Indirectness Imprecision Other considerations ethambutol no ethambutol  Aure/relapse vs treatment success - Strain susceptible to Ethambutol  Observational studies not serious not serious serious serious not serious not serious serious none 123/1163 (10.6%) 50/411 (12.2%)  Observational studies not serious not serious serious not seriou	Study design Risk of bias Inconsistency Indirectness Imprecision Other considerations ethambutol no ethambutol (95% CI)  underprelapse vs treatment success - Strain susceptible to Ethambutol  observational studies not serious not serious not serious serious serious not serious serious serious - none 123/1163 (10.6%) 50/411 (12.2%) (0.10 to 0.99)  underprelapse vs treatment success - Strain susceptible to Ethambutol  observational not serious not serious not serious serious serious not serious serious - none 1205/1245 (16.5%) 55/416 (13.2%) (0.1 to 1.7)  underprelapse vs treatment success - Strain resistant to Ethambutol  observational not serious not serious not serious serious serious serious - none 122/424 (15.1%) 228/1974 (11.6%) aoR 1.3 (0.7 to 2.3)  underprelapse vs treatment success - Strain resistant to Ethambutol  observational not serious not serious not serious serious serious serious - none 122/482 (25.3%) 347/2093 (16.6%) aoR 1.3 (0.7 to 2.3)  underprelapse vs treatment success - Strain resistant to Ethambutol  observational not serious not serious not serious serio	Study design Risk of bias Inconsistency Indirectness Imprecision Other considerations ethambutol no ethambutol (95% C)  Unreletapse us treatment success - Strain susceptible to Ethambutol  doservational studies not serious not serious not serious surfer particular process of the process of	Study design Risk of bias inconsistency indirectness imprecision Other considerations ethamburol no ethamburol (95% C)

CI: Confidence interval

- a. Small numbers
- b. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- c. Simulated I squared from one-step IPD exceeds 50%
- d. Pooled incidence of adverse events of random effect in meta-analysis

Author(s): Investigators of Trial 213 of delamanid use in adults reported for the WHO treatment guidelines for multidrug- and rifampicin-resistant tuberculosis, 2018 update.

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Phase III trial on delamanid (17 centres in 7 countries; as reported in October 2017)

Bibliography: Otsuka Pharmaceutical Development &, Commercialization, Inc. Clinical Study Report. A Phase 3, Multicenter, Randomized, Double-blind, Placebo-controlled, Parallel Group Trial to Evaluate the Safety and Efficacy of Delamanid (OPC-67683) Administered Orally as 200 mg Total Daily Dose for Six Months in Patients With Pulmonary Sputum Culture-positive, Multidrug-resistant Tuberculosis (Protocol No. 242-09-213). Rockville (MD), United States; 7 June 2017.

#### 7. Delamanid (adults)

			Certainty a	assessment			N₂ofp	patients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	delamanid	no delamanid	Relative (95% CI)	Absolute (95% CI)		
Treatment su	ccess at 30 month	s (follow up: range 6	months to 12 months;	assessed with: Trial 2	213; ITT population; N	MGIT) <sup>a</sup>						
1	randomised trials <sup>b</sup>	not serious °	not serious <sup>d</sup>	not serious <sup>e</sup>	serious ¹	none	263/341 (77.1%)	132/170 (77.6%)	RR 0.993 (0.899 to 1.097)	<b>5 fewer per</b> <b>1,000</b> (from 75 more to 78 fewer)	MODERATE	CRITICAL
Mortality at 3	0 months (follow u	p: range 6 months to	12 months; assessed	with: Trial 213; ITT po	opulation) <sup>g</sup>		,					
1	randomised trials <sup>b</sup>	not serious <sup>c</sup>	not serious <sup>d</sup>	not serious <sup>e</sup>	serious <sup>1</sup>	none	18/341 (5.3%)	8/170 (4.7%)	<b>RR 1.122</b> (0.498 to 2.527)	6 more per 1,000 (from 24 fewer to 72 more)	MODERATE	CRITICAL
Serious Adve	rse Events (follow	up: range 6 months t	o 12 months; assesse	ed with: Trial 213; ITT	population)		1					1
1	randomised trials	not serious	not serious	not serious <sup>e</sup>	serious <sup>th</sup>	none	89/341 (26.1%) <sup>1</sup>	47/170 (27.6%) <sup>1</sup>	RR 0.944 (0.698 to 1.276)	15 fewer per 1,000 (from 76 more to 83 fewer)	MODERATE	CRITICAL
QTcF interva	prolongation >60r	ns from baseline on e	electrocardiogram ove	r 30 months (follow u	o: range 6 months to	12 months; assessed with: Trial 2:	13; ITT population) <sup>№</sup>	l				
1	randomised trials	not serious	not serious	not serious <sup>e</sup>	serious <sup>1</sup>	none	35/341 (10.3%)	12/170 (7.1%)	<b>RR 1.454</b> (0.775 to 2.728) <sup>m</sup>	32 more per 1,000 (from 16 fewer to 122 more)	₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩₩	CRITICAL
QTcF interva	>500ms (new ons	set) on electrocardiog	ram over 30 months (1	follow up: range 6 mo	nths to 12 months; as	ssessed with: Trial 213; ITT popula	ation) <sup>kn</sup>					<u> </u>
1	randomised trials	not serious	not serious	not serious <sup>e</sup>	serious <sup>1</sup>	none	7/341 (2.1%)	2/170 (1.2%)	OR 1.761 (0.362 to 8.568) <sup>m</sup>	9 more per 1,000 (from 7 fewer to 81 more) <sup>m</sup>	⊕⊕⊕ MODERATE	CRITICAL

			Certainty a	assessment			Nº of p	atients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	delamanid	no delamanid	Relative (95% CI)	Absolute (95% CI)		
Acquired resi	stance to delaman	id up to 26 weeks (as	ssessed with: Trial 213	B; ITT population; MG	T or solid media cultu	re)°						
1	randomised trials	serious <sup>p</sup>	not serious	not serious <sup>e</sup>	serious <sup>1</sup>	none	4/341 (1.2%) <sup>q</sup>	0/170 (0.0%)	not estimable		⊕⊕⊖ Low	CRITICAL
Time to sputu	ım culture convers	ion by 6 months (asse	essed with: Trial 213;	MITT population; MG	IT)							
1	randomised trials	not serious	not serious	serious <sup>e,r</sup>	not serious	none	198/226 (87.6%) <sup>s</sup>	87/101 (86.1%) <sup>s</sup>	HR 1.17 (0.91 to 1.51)	<b>40 more per 1,000</b> (from 27 fewer to 88 more)	MODERATE	IMPORTANT
Sputum cultu	re conversion at 2	months (assessed wi	ith: Trial 213; MITT po	pulation; MGIT) <sup>t</sup>	ļ.							
1	randomised trials	not serious	not serious	serious <sup>r</sup>	not serious	none	132/226 (58.4%)	54/101 (53.5%)	RR 1.096 (0.889 to 1.352)	<b>51 more per 1,000</b> (from 59 fewer to 188 more)	MODERATE MODERATE	IMPORTANT
Sputum cultu	re conversion at 6	months (assessed wi	ith: Trial 213; MITT po	pulation; MGIT) <sup>u</sup>								
1	randomised trials	not serious	not serious	serious <sup>r</sup>	not serious	none	198/226 (87.6%)	87/101 (86.1%)	<b>RR 1.017</b> (0.927 to 1.115)	15 more per 1,000 (from 63 fewer to 99 more)	⊕⊕⊕⊖ MODERATE	IMPORTANT

CI: Confidence interval; RR: Risk ratio; OR: Odds ratio; HR: Hazard Ratio

- a. Treatment success: defined as per Trial 213, sustained sputum culture conversion (SCC) by week 26, completing the trial out to month 30 visit with SCC and staying alive at the last contact for follow-up per trial (CT-5.10.1.2 in page 2042 of Clinical Study Report PROTOCOL 242-09-213). Results with solid media obtain lead to similar estimates of effects RR=1.008 (95%CL: 0.916, 1.110) (CT-5.10.2.2 of PROTOCOL 242-09-213; page 2051). RRs using results at different time points (18 and 24 months) are very similar to the value shown here (CT-CT-5.8.1.2 in page 2026 of Clinical Study Report PROTOCOL 242-09-213).
- b. Final outcomes such as treatment success and death were secondary outcomes of Trial 2013, which was primarily powered to assess time to SCC
- c. Patients without SCC by 6 months could have had a different treatment continuation. Violations of masking were judged to be unlikely as based on the distribution of patient characteristics in the delamanid and placebo arms.
- d. Only one study undertaken
- e. XDR-TB cases were excluded and many patients received combination chemotherapy in the 30 days before treatment (see Table CT-4.4.1 in page 1840 of Clinical Study Report PROTOCOL 242-09-213); otherwise the study population relates closely to adult patients in whom this medicine is likely to be indicated. The Expert Panel did not downgrade for the exclusion of children in the trial, because it considered that any recommendation for individuals under 18 years of age would be based on the same principles used to develop the interim policy for 6-17 year olds in 2016, i.e. that efficacy in paediatric TB patients can be inferred from adult PK/PD data coupled with paediatric PK data can inform about what drug dosages in children will achieve adult PK targets, so the extrapolation of delamanid recommendations from adults to children was reasonable (see WHO guidelines at apps.who.int/inis/bistream/10665/250614/19/789241549899-eng.pdf)
- f. Downgraded to serious for non primary outcomes (i.e. not the endpoints used for sample size calculation); heterogeneity could not be assessed properly, despite that this MDR-TB treatment trial was of moderate-size and conducted in 17 sites in 7 countries.
- g. All-cause mortality assessed during the whole 30 months (see CT-5.11.1 in page 2055 of Clinical Study Report PROTOCOL 242-09-213)
- h. Trials of this type are not meant to provide an exhaustive understanding of the safety profile of a medicine, particularly for harms which are rare and that are best assessed with post-marketing surveillance

- i. As per serious Treatment Emergent Adverse Events (see CT-8.1.1 in page 2227 of Clinical Study Report PROTOCOL 242-09-213)
- i. Calculated using the values in Table CT-8.1.1
- k. As per Table CT-11.4.1 of PROTOCOL 242-09-213; page 5002
- I. 8 of 35 cases with QTcF prolongation in the delamanid group and 5/12 of the placebo group happened after 6 months.
- m. Calculated manually using values in Table CT-11.4.1 of PROTOCOL 242-09-213; page 5002
- n. 3 of 7 cases with QTcF prolongation in the delamanid group and 2/2 of the placebo group happened after 6 months.
- o. Delamanid susceptibility is not available for all of the 511 subjects (in page 379 of Clinical Study Report PROTOCOL 242-09-213 a denominator of 502 is reported). The calculation is done for the 341 patients exposed to delamanid (retaining in the denominator the 2 cases found to be delamanid resistant at baseline). The DST methodology used only assessed resistance at baseline and week 26 and not on the last isolate, and therefore this is likely to be a minimum estimate
- p. DST was not systematically done on the last positive culture for patients receiving delamanid (testing was done at 26 weeks and acquired resistance may thus have been underestimated). Resistance only emerged in patients on 3-drug regimens.
- q. 95% binomial confidence intervals (exact): 0.3-3%
- r. The culture conversion outcomes were surrogate endpoints for population-important outcomes (cure or failure) and therefore downgraded by one.
- s. Number of cases converting by 6 months (as per CT-5.1.1.1 in page 1954 of Clinical Study Report PROTOCOL 242-09-213); the primary analysis used a 2-sided stratified Peto-Peto modification of Gehan's Wilcoxon rank sum test to compare the survival distribution curves of time to SCC as per regulatory requirement. A protocol-specified sensitivity analysis using "last observation carried forward" (LOCF) resulted in a HR of 1.24 (95%Cl 0.96-1.6) and a HR of 1.33 (95%Cl 1.03-1.74) using the "bookended" method (ST-16.1 in page 6024 of Clinical Study Report PROTOCOL 242-09-213). Median times to SCC were: (i) ITT analysis 51 days vs 57 days (P-value 0.056); (ii) LOCF 44 days vs 57 days (P-value 0.0281); (iii) "bookended" 51 days vs 64 days (P-value0.0052). Other analysis for sustained conversion shows a reduced effect.
- t. RR was estimated using PROC FREQ (SAS) and the Cochran-Mantel-Haenszel general association test (see CT-5.2.1.1 in page 1964 of Clinical Study Report PROTOCOL 242-09-213). Using LOCF (CT-5.2.1.2), the RR was 1.069 (95%CI: 0.875-1.305) and the "bookended" (ST-16.3) RR was 1.172 (95%CI: 0.946-1.453)
- u. RR was estimated using PROC FREQ (SAS) and the Cochran-Mantel-Haenszel general association test (see CT-5.3.1.1 in page 1972 of Clinical Study Report PROTOCOL 242-09-213). Using LOCF (CT-5.3.1.2), the RR 1.034 (95%CI: 0.941-1.137) and the "bookended" (ST-16.2) RR was 1.070 (95%CI: 0.950-1.204).

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

#### 8. Pyrazinamide

			Certainty a	assessment			Nº of p	atients	Effec	t	Certainty	Importance
N₂ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	pyrazinamide	no pyrazinamide	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	lure/relapse vs. tre	eatment success - Str	ains susceptible to PZ	'A								
53	observational studies	not serious	not serious	not serious	not serious	none	98/1248 (7.9%)	4/100 (4.0%)	<b>aOR 2.7</b> (0.7 to 10.9)	7 more per 100 (from 2 fewer to 16 more)	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - S	Strains susceptible to	PZA									
53	observational studies	not serious	not serious	not serious	not serious	none	122/1272 (9.6%)	7/103 (6.8%)	<b>aOR 1.2</b> (0.1 to 15.7)	1 more per 100 (from 8 fewer to 10 more)	⊕⊕⊖⊖ Low	CRITICAL
Treatment fai	lure/relapse vs. tre	eatment success - Str	ains resistant to PZA									
53	observational studies	not serious	not serious	not serious	not serious	none	89/790 (11.3%)	31/434 (7.1%)	<b>aOR 1.9</b> (0.9 to 3.8)	4 more per 100 (from 1 fewer to 10 more)	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - S	Strains resistant to PZ	'A	L	<u> </u>	l	l					
53	observational studies	not serious	not serious	not serious	not serious	none	119/820 (14.5%)	49/452 (10.8%)	<b>aOR 1.2</b> (0.6 to 2.4)	1 more per 100 (from 4 fewer to 6 more)	⊕⊕⊖⊖ Low	CRITICAL
Sensitivity An	alysis - Excluding	patients who received	d LZD, BDQ, CFZ or C	Carbapenems (No res	triction on number of	effective drugs) - Treatment failure	e/relapse vs. treatment si	uccess				
53	observational studies	not serious	not serious	not serious	not serious	none	100/1323 (7.6%)	10/135 (7.4%)	<b>aOR 0.7</b> (0.1 to 8.7)	0 fewer per 100 (from 8 fewer to 7 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Sensitivity An	alysis - Excluding	L patients who received	LZD, BDQ, CFZ or C	L Carbapenems (No res	triction on number of	I effective drugs) - Death vs. treatm	ent success					l

	Certainty assessment							№ of patients		Effect		Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	pyrazinamide	no pyrazinamide	Relative (95% CI)	Absolute (95% CI)		
53	observational studies	not serious	not serious	not serious	not serious	none	131/1354 (9.7%)	18/143 (12.6%)	<b>aOR 0.4</b> (0.0 to 7.5)	5 fewer per 100 (from 13 fewer to 4 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Adverse ever	Adverse events (drug stopped permanently)											
31	observational studies	not serious	serious <sup>b</sup>	not serious	not serious	none	394/4975 (7.9%)		not estimable	5 more per 100 (from 3 more to 8 more) °	⊕⊖⊖⊖ VERY LOW	IMPORTANT

CI: Confidence interval

- a. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- b. Simulated I squared from one-step IPD exceeds 50%
- c. Pooled incidence of adverse events of random effect in meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

#### 9. Imipenem-cilastatin / Meropemen

			Certainty a	ssessment			№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	carbapenems (imipenem- cilastatin / meropenem)	no carbapenems	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	lure/relapse vs. tre	atment success - Str	ain susceptible to Car	bapenems						'		
53	observational studies	serious <sup>a</sup>	not serious	not serious	serious <sup>b</sup>	none	17/206 (8.3%)	675/5798 (11.6%)	<b>aOR 0.4</b> (0.2 to 0.7)	11 fewer per 100 (from 19 fewer to 3 fewer) °	⊕⊖⊖⊖ VERY LOW	CRITICAL
Death vs. trea	atment success - S	train susceptible to C	Carbapenems									•
53	observational studies	serious <sup>a</sup>	serious <sup>d</sup>	not serious	serious <sup>b</sup>	none	15/204 (7.4%)	1489/6612 (22.5%)	<b>aOR 0.2</b> (0.1 to 0.5)	18 fewer per 100 (from 27 fewer to 8 fewer) °	⊕⊖⊖ VERY LOW	CRITICAL
Sensitivity An	alysis - Patients w	th XDR (no restriction	n on number of effecti	ve drugs) - Treatment	t failure/relapse vs. tre	eatment success						
53	observational studies	serious <sup>a</sup>	not serious	not serious	serious <sup>b</sup>	none	12/121 (9.9%)	238/930 (25.6%)	<b>aOR 0.5</b> (0.2 to 1.2)	8 fewer per 100 (from 19 fewer to 2 more) °	VERY LOW	CRITICAL
Sensitivity An	alysis - Patients w	th XDR (no restriction	n on number of effecti	ve drugs) - Death vs.	treatment success	L						
53	observational studies	serious <sup>a</sup>	not serious	not serious	serious <sup>b</sup>	none	20/129 (15.5%)	547/1239 (44.1%)	<b>aOR 0.50</b> (0.20 to 1.05)	11 fewer per 100 (from 24 fewer to 1 more) °	⊕⊖⊖ VERY LOW	CRITICAL
Adverse ever	nts (meropenem dr	ug permanently stop	ped)									
7	observational studies	serious <sup>e</sup>	not serious	not serious	not serious	none	9/158 (5.7%)		not estimable	5 more per 100 (from 1 more to 20 more) <sup>1</sup>	⊕⊖⊖⊖ VERY LOW	IMPORTANT

CI: Confidence interval

- a. 95% patients who received carbapenems also received at least one other newer drug (i.e. bedaquiline, linezolid or clofazimine). A sensitivity analysis excluding patients who received other newer drugs could not be performed.
- b. Randomization method unclear
- c. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- d. Small numbers
- e. Two out of four studies that reported adverse events for carbapenems only reported adverse events for selected drugs.
- f. Pooled incidence of adverse events of random effect in meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

#### 9. Clavulanic acid

			Certainty a	ssessment			№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	amoxicillin- clavulanate	no amoxicillin- clavulanate	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	Treatment failure/relapse vs. treatment success - Strain susceptible to Amoxicilin-clavulanate											
53	observational studies	not serious	not serious	not serious	not serious	none	109/492 (22.2%)	383/3500 (10.9%)	<b>aOR 1.70</b> (1.02 to 3.00)	8 more per 100 (from 1 more to 15 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - S	Strain susceptible to A	moxicilin-clavulanate									
53	observational studies	not serious	not serious	not serious	not serious	strong association	151/534 (28.3%)	888/4005 (22.2%)	<b>aOR 2.2</b> (1.3 to 3.6)	12 more per 100 (from 5 more to 19 more) <sup>a</sup>	⊕⊕⊕⊖ MODERATE	CRITICAL
Adverse ever	Adverse events (drug stopped permanently)											
20	observational studies	not serious	not serious	not serious	not serious	none	18/654 (2.8%)		not estimable	3 more per 100 (from 1 more to 5 more) <sup>b</sup>	⊕⊕⊖⊖ Low	IMPORTANT

CI: Confidence interval

### **Explanations**

a. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records

b. Pooled incidence of adverse events from random effects meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

#### 10. Amikacin

Certainty assessment								№ of patients		Effect		Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	amikacin	no injectable agent or different one	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	lure/relapse vs. tre	eatment success - Am	ikacin vs NO INJECT	ABLE			<b>'</b>					'
53	observational studies	not serious	not serious	not serious	not serious	none	37/635 (5.8%)	64/669 (9.6%)	<b>aOR 0.3</b> (0.1 to 0.8)	8 fewer per 100 (from 15 fewer to 1 fewer) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - A	mikacin vs NO INJE	CTABLE				<u>'</u>			, ,		-
53	observational studies	not serious	not serious	not serious	not serious	none	129/727 (17.7%)	157/762 (20.6%)	<b>aOR 0.7</b> (0.4 to 1.2)	6 fewer per 100 (from 16 fewer to 4 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Treatment fai	lure/relapse vs. tre	eatment success - Am	ikacin (S) vs Capreon	nycin (S)			<b>'</b>					•
53	observational studies	not serious	not serious	not serious	not serious	none	37/635 (5.8%)	93/777 (12.0%)	<b>aOR 0.3</b> (0.1 to 0.6)	7 fewer per 100 (from 13 fewer to 1 fewer) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - A	mikacin (S) vs Capre	eomycin (S)				·	l		<u> </u>		1
53	observational studies	not serious	not serious	not serious	not serious	none	129/727 (17.7%)	142/826 (17.2%)	<b>aOR 1.0</b> (0.6 to 1.7)	1 fewer per 100 (from 8 fewer to 7 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Treatment fai	ilure/relapse vs. tre	eatment success - Am	ikacin (S) vs Kanamy	cin (S)	L							
53	observational studies	not serious	not serious	not serious	not serious	none	37/635 (5.8%)	308/2946 (10.5%)	<b>aOR 0.4</b> (0.2 to 0.8)	7 fewer per 100 (from 11 fewer to 3 fewer) <sup>a</sup>	ФфСС	CRITICAL
eath vs. trea	atment success - A	mikacin (S) vs Kanar	mycin (S)									

			Certainty a	ssessment			N₂ofp	patients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	amikacin	no injectable agent or different one	Relative (95% CI)	Absolute (95% CI)		
53	observational studies	not serious	not serious	not serious	not serious	none	129/727 (17.7%)	631/3269 (19.3%)	<b>aOR 0.9</b> (0.7 to 1.3)	1 fewer per 100 (from 6 fewer to 4 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Sensitivity Ar	nalysis - Treatment	failure/relapse vs. tre	eatment success - Ami	kacin used (all) vs NC	) Injectable - in XI	DR						
53	observational studies	not serious	not serious	not serious	serious <sup>b</sup>	none	12/96 (12.5%)	43/339 (12.7%)	<b>aOR 1.0</b> (0.4 to 3.0)	0 fewer per 100 (from 11 fewer to 11 more) <sup>a</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Sensitivity Ar	nalysis - Death vs. t	treatment success - A	mikacin used (all) vs	NO INJECTABLE - in	XDR							
53	observational studies	not serious	not serious	not serious	serious <sup>b</sup>	none	20/104 (19.2%)	93/389 (23.9%)	<b>aOR 0.8</b> (0.3 to 2.3)	3 fewer per 100 (from 18 fewer to 12 more) <sup>a</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Adverse ever	nts (amikacin drug	stopped permanently	)									
18	observational studies	not serious	serious <sup>c</sup>	not serious	not serious	none	214/4023 (5.3%)		not estimable	7 more per 100 (from 5 more to 12 more) <sup>d</sup>	⊕⊖⊖ VERY LOW	IMPORTANT

CI: Confidence interval

- a. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- b. Small numbers
- c. Simulated I squared from one-step IPD exceeds 50%
- d. Pooled incidence of adverse events of random effect in meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

### 10a. Streptomycin

			Certainty a	ssessment			N₂ of p	patients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	streptomycin	no injectable agent or different one	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	ilure/relapse vs. tre	eatment success - Str	eptomycin vs NO INJE	ECTABLE								
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	8/226 (3.5%)	64/669 (9.6%)	<b>aOR 0.5</b> (0.1 to 2.1)	4 fewer per 100 (from 11 fewer to 4 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Death vs. trea	atment success - S	Streptomycin vs NO IN	NJECTABLE									
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	20/238 (8.4%)	157/762 (20.6%)	<b>aOR 0.1</b> (0.0 to 0.4)	26 fewer per 100 (from 40 fewer to 12 fewer) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Adverse ever	nts (streptomycin d	rug stopped permane	ently)									
16	observational studies	not serious	serious <sup>c</sup>	not serious	serious <sup>a</sup>	none	33/1145 (2.9%)		not estimable	3 more per 100 (from 1 more to 7 more) <sup>d</sup>	⊕⊖⊖⊖ VERY LOW	IMPORTANT

CI: Confidence interval

- a. Small numbers
- b. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- c. Simulated I squared from one-step IPD exceeds 50%
- d. Pooled incidence of adverse events of random effect in meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

### 10b. Kanamycin

			Certainty a	ssessment			Nº of p	patients	Effec	1	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	kanamycin	no injectable agent or different one	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	lure/relapse vs. tre	atment success - Kar	namycin vs NO INJEC	TABLE								
53	observational studies	not serious	not serious	not serious	not serious	none	308/2946 (10.5%)	64/669 (9.6%)	<b>aOR 1.90</b> (1.01 to 3.40)	7 more per 100 (from 1 fewer to 15 more) <sup>b</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - K	anamycin vs NO INJ	ECTABLE									
53	observational studies	not serious	not serious	not serious	not serious	none	631/3269 (19.3%)	157/762 (20.6%)	<b>aOR 1.1</b> (0.5 to 2.1)	2 more per 100 (from 10 fewer to 13 more) <sup>b</sup>	⊕⊕⊖⊖ Low	CRITICAL
Treatment fai	lure/relapse vs. tre	atment success - Am	ikacin (S) vs Kanamy	cin (S)								
53	observational studies	not serious	not serious	not serious	not serious	none	37/635 (5.8%)	308/2946 (10.5%)	<b>aOR 0.4</b> (0.2 to 0.8)	7 fewer per 100 (from 11 fewer to 3 fewer) <sup>b</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - A	mikacin (S) vs Kanar	nycin (S)									
53	observational studies	not serious	not serious	not serious	not serious	none	129/727 (17.7%)	631/3269 (19.3%)	<b>aOR 0.9</b> (0.7 to 1.3)	1 fewer per 100 (from 6 fewer to 4 more) <sup>b</sup>	⊕⊕⊖⊖ Low	CRITICAL
Sensitivity An	alysis - Treatment	failure/relapse vs. tre	atment success - Kan	amycin used (all) vs i	NO INJECTABLE - in	XDR						
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	36/128 (28.1%)	43/339 (12.7%)	<b>aOR 2.2</b> (1.1 to 4.7)	13 more per 100 (from 2 more to 24 more) <sup>b</sup>	VERY LOW	CRITICAL

Sensitivity Ar	nalysis - Death vs.	treatment success - k	Canamycin used (all) v	'S NO INJECTABLE -	in XDR							
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	37/129 (28.7%)	93/389 (23.9%)	<b>aOR 1.4</b> (0.6 to 3.0)	6 more per 100 (from 8 fewer to 20 more) <sup>b</sup>	⊕⊖⊖⊖ VERY LOW	CRITICAL
Adverse ever	nts (kanamycin dru	g stopped permanen	tly)									
25	observational studies	not serious	serious <sup>c</sup>	not serious	not serious	none	268/1995 (13.4%)		not estimable	8 more per 100 (from 5 more to 12 more) <sup>d</sup>	⊕⊖⊖⊖ VERY LOW	IMPORTANT

CI: Confidence interval

- a. Small numbers
- b. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- c. Simulated I squared from one-step IPD exceeds 50%
- d. Pooled incidence of adverse events of random effect in meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

### 10c. Capreomycin

			Certainty a	ssessment			N₂ of p	patients	Effec	t	Certainty	Importance
Nº of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	capreomycin	no injectable agent or different one	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	lure/relapse vs. tre	atment success - Ca	preomycin vs NO INJ	ECTABLE								
53	observational studies	not serious	not serious	not serious	not serious	none	93/777 (12.0%)	64/669 (9.6%)	<b>aOR 2.0</b> (1.1 to 3.5)	7 more per 100 (from 0 more to 14 more) <sup>b</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - C	Capreomycin vs NO II	NJECTABLE									
53	observational studies	not serious	not serious	not serious	not serious	none	142/826 (17.2%)	157/762 (20.6%)	<b>aOR 1.4</b> (0.7 to 2.8)	7 more per 100 (from 6 fewer to 19 more) <sup>b</sup>	⊕⊕⊖⊖ Low	CRITICAL
Treatment fai	lure/relapse vs. tre	atment success - Am	ikacin (S) vs Capreon	nycin (S)								
53	observational studies	not serious	not serious	not serious	not serious	none	37/635 (5.8%)	93/777 (12.0%)	<b>aOR 0.3</b> (0.1 to 0.6)	7 fewer per 100 (from 13 fewer to 1 fewer) <sup>b</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. trea	atment success - A	mikacin (S) vs Capre	omycin (S)							'		
53	observational studies	not serious	not serious	not serious	not serious	none	129/727 (17.7%)	142/826 (17.2%)	<b>aOR 1.0</b> (0.6 to 1.7)	1 fewer per 100 (from 8 fewer to 7 more) <sup>b</sup>	⊕⊕⊖⊖ Low	CRITICAL
Sensitivity An	alysis - Treatment	failure/relapse vs. tre	atment success - Cap	oreomycin used (all) v	s NO INJECTABLE -	in XDR						
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	136/398 (34.2%)	43/339 (12.7%)	<b>aOR 1.6</b> (0.9 to 2.8)	6 more per 100 (from 2 fewer to 14 more) <sup>b</sup>	₩ VERY LOW	CRITICAL

Sensitivity Ar	nalysis - Death vs.	treatment success - C	Capreomycin used (all	) vs NO INJECTABLE	- in XDR							
Adverse even	observational studies	not serious rug stopped permane	not serious	not serious	serious <sup>a</sup>	none	368/630 (58.4%)	93/389 (23.9%)	<b>aOR 2.0</b> (1.2 to 3.3)	14 more per 100 (from 2 more to 25 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
25	observational studies	not serious	not serious	not serious	not serious	none	150/1861 (8.1%)		not estimable	8 more per 100 (from 6 more to 10 more) °	⊕⊕⊖⊖ Low	IMPORTANT

CI: Confidence interval

- a. Small numbers
- b. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- c. Pooled incidence of adverse events of random effect in meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

### 11. Ethionamide / Prothionamide

			Certainty a	assessment			N₂ of p	atients	Effec	ı	Certainty	Importance
Nº of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	prothionamide or ethionamide	no prothionamide or ethionamide	Relative (95% CI)	Absolute (95% CI)		
Treatment fai	ilure/relapse vs. tre	eatment success - Str	ain susceptible to Pro	thionamide or Ethiona	ımide							
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	327/2582 (12.7%)	14/147 (9.5%)	<b>aOR 1.6</b> (0.5 to 5.5)	3 more per 100 (from 5 fewer to 12 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Death vs. trea	atment success - S	Strain susceptible to F	rothionamide or Ethic	onamide								
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	495/2750 (18.0%)	12/145 (8.3%)	<b>aOR 2.0</b> (0.8 to 5.3)	7 more per 100 (from 3 fewer to 17 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Treatment fai	ilure/relapse vs. tre	eatment success - Str	ain resistant to Prothic	onamide or Ethionami	de							
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	35/344 (10.2%)	31/265 (11.7%)	<b>aOR 1.3</b> (0.5 to 3.1)	3 more per 100 (from 7 fewer to 14 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Death vs. trea	atment success - S	Strain resistant to Pro	hionamide or Ethiona	mide	·							
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	61/370 (16.5%)	22/256 (8.6%)	<b>aOR 3.0</b> (1.5 to 6.0)	12 more per 100 (from 2 more to 22 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Sensitivity Ar	nalysis - Excluding	patients who received	LZD, BDQ, CFZ or C	Carbapenems (no rest	riction on number of e	effective drugs) - Treatment failure	/relapse vs. treatment su	iccess				
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	340/2562 (13.3%)	14/136 (10.3%)	<b>aOR 0.7</b> (0.1 to 6.2)	1 fewer per 100 (from 10 fewer to 8 more) b	VERY LOW	CRITICAL

Sensitivity	Analysis - Excluding	patients who receive	d LZD, BDQ, CFZ or 0	Carbapenems (no rest	triction on number of e	effective drugs) - Death vs. treatm	ent success					
53	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	529/2751 (19.2%)	36/158 (22.8%)	<b>aOR 0.4</b> (0.0 to 4.9)	8 fewer per 100 (from 18 fewer to 1 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL
Adverse e	vents (drugs stopped	permanently)										
35	observational studies	not serious	serious <sup>c</sup>	not serious	serious <sup>a</sup>	none	357/4537 (7.9%)		not estimable	6 more per 100 (from 3 more to 9 more) <sup>d</sup>	⊕⊖⊖ VERY LOW	IMPORTANT

CI: Confidence interval

- a. Small numbers
- b. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- c. Simulated I squared from one-step IPD exceeds 50%
- d. Pooled incidence of adverse events of random effect in meta-analysis

Date: 16 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The main IPD-MA dataset consisted of 13,104 records from 53 studies in 40 countries.

### 12. p-aminosalicylic acid (PAS)

		Certainty a	assessment			N₂ of p	oatients	Effec	t	Certainty	Importance
Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	p-aminosalicylic acid (PAS)	no p-aminosalicylic acid (PAS)	Relative (95% CI)	Absolute (95% CI)		
lure/relapse vs. tre	atment success - Str	ain susceptible to PAS	5								
observational studies	not serious	not serious	not serious	not serious	none	184/1564 (11.8%)	36/580 (6.2%)	<b>aOR 3.1</b> (1.1 to 8.9)	5 more per 100 (from 1 more to 9 more) <sup>a</sup>	⊕⊕⊖ Low	CRITICAL
atment success - S	strain susceptible to F	PAS			•						
observational studies	not serious	not serious	not serious	not serious	none	229/1609 (14.2%)	76/620 (12.3%)	<b>aOR 1.0</b> (0.6 to 1.6)	0 fewer per 100 (from 5 fewer to 5 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
ilure/relapse vs. tre	atment success - Str	ain resistant to PAS	L	L		ļ					
observational studies	not serious	not serious	not serious	serious <sup>b</sup>	none	13/96 (13.5%)	4/114 (3.5%)	<b>aOR 3.2</b> (0.9 to 11.5)	7 more per 100 (from 1 fewer to 15 more) <sup>a</sup>	₩ CERY LOW	CRITICAL
atment success - S	Strain resistant to PAS	5			L						
observational studies	not serious	not serious	not serious	serious <sup>b</sup>	none	14/97 (14.4%)	6/116 (5.2%)	<b>aOR 2.2</b> (0.7 to 6.8)	6 more per 100 (from 3 fewer to 14 more) <sup>a</sup>	⊕⊖⊖ VERY LOW	CRITICAL
nts (drug stopped o	completely)				<u>'</u>						
observational studies	not serious	serious <sup>c</sup>	not serious	not serious	none	506/2888 (17.5%)		not estimable	10 more per 100 (from 6 more to 15 more) <sup>d</sup>	₩ CONTRACTOR VERY LOW	IMPORTANT
	observational studies  observational studies	observational studies not serious  observational studies not serious  atment success - Strain susceptible to F  observational studies not serious  lure/relapse vs. treatment success - Str  observational studies not serious  atment success - Strain resistant to PAS  observational studies not serious  observational studies not serious  observational not serious  observational not serious	Study design Risk of bias Inconsistency  Illure/relapse vs. treatment success - Strain susceptible to PAS  observational studies not serious not serious  observational not serious not serious  observational studies not serious not serious  observational not serious not serious  observational not serious not serious  observational not serious serious  observational not serious serious	Ilure/relapse vs. treatment success - Strain susceptible to PAS	Study design   Risk of bias   Inconsistency   Indirectness   Imprecision	Study design   Risk of bias   Inconsistency   Indirectness   Imprecision   Other considerations	Study design   Risk of bias   Inconsistency   Indirectness   Imprecision   Other considerations   P-aminosalicylic acid (PAS)	Study design Risk of bias Inconsistency Indirectness Imprecision Other considerations P-aminosalicylic acid (PAS) no p-aminosalicylic acid (PAS)  Iurefrelapse vs. treatment success - Strain susceptible to PAS  observational success - Strain susceptible to PAS  observational not serious not serious not serious not serious not serious none 184/1564 (11.8%) 36/580 (6.2%)  urefrelapse vs. treatment success - Strain susceptible to PAS  observational success - Strain resistant to PAS  observational not serious not serious not serious serious none 13/96 (13.5%) 4/114 (3.5%)  observational not serious not serious not serious serious none 14/97 (14.4%) 6/116 (5.2%)  observational not serious not serious not serious serious serious none 14/97 (14.4%) 6/116 (5.2%)  observational not serious not serious not serious not serious serious serious serious serious success - Strain resistant to PAS	Study design Risk of bias Inconsistency Indirectness Imprecision Other considerations P-aminosalicylic acid (PAS) Relative (95% CI)  underprelapse vs. treatment success - Strain susceptible to PAS  observational studies not serious not serious not serious not serious not serious none 184/1564 (11.8%) 36/580 (6.2%) aOR 3.1 (1.1 to 8.9)  observational studies not serious not serious not serious not serious none 229/1609 (14.2%) 76/620 (12.3%) (0.6 to 1.6)  underprelapse vs. treatment success - Strain resistant to PAS  observational not serious not serious not serious serious sudies not serious not serious not serious serious serious serious not serious serious serious serious not ser	Study design Risk of bias Inconsistency Indirectness Imprecision Other considerations P-aminosalicylic acid (PAS): Relative acid (PAS):	Study design Risk of bias Inconsistency Indirectness Imprecision Other considerations P-aminosalicytic acid (PAS) no p-aminosalicytic (95% CI) (15%

CI: Confidence interval

- a. Risk differences and confidence interval from propensity score matched regression meta-analysis of individual patient records
- b. Small numbers
- c. Simulated I squared from one-step IPD exceeds 50%
- d. Pooled incidence of adverse events of random effect in meta-analysis

Author(s): Viney K et al. Literature review for outcomes of MDR-TB patient treated with regimens including perchlozone

Date: 30 July 2018

Question: PICO 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?

Setting: Health facilities, Russian Federation. (Pavlova M, et al. Efficacy of complex therapy and possible adverse reactions of respiratory apparatus multi-drug resistance tuberculosis treatment. Saint Petersburg, Russian Federation Saint Petersburg State University 2014; Pavlova M, et al. Confirmation of preclinical studies data on Perchlorate (Perchlozon) in clinic. International Journal of Current Research 2015;7(4):14501-5; Chernokhaeva I. The effectiveness tioureidoiminomethylpyridiunium perchlorate in the treatment of multidrug-resistant tuberculosis. Saint Petersburg, Russian Federation Saint Petersburg State University 2017.; Belyaeva E. Features of the course and effectiveness of therapy for extensively drug resistant pulmonary tuberculosis Saint Petersburg, Russian Federation Saint Petersburg State University 2018.; Yablonskii P, et al. Preclinical and clinical trials of the new tuberculosis drug perchlozon. Therapeutic Archive. 2016;88(3):111-5))

#### 13. Perchlozone

			Certainty a	assessment			N₂ of p	atients	Effec	t	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	perchlozone	no perchlozone	Relative (95% CI)	Absolute (95% CI)		
Six to eight n	nonth culture conve	ersion (follow up: mea	an 6.4 months)									
4	randomised trials	serious <sup>a</sup>	not serious <sup>b</sup>	serious °	serious <sup>d</sup>	none	100/144 (69.4%)	76/129 (58.9%)	<b>OR 1.64</b> (0.76 to 3.52)	11 more per 100 (from 7 fewer to 25 more)	⊕⊖⊖ VERY LOW	IMPORTANT
Adverse ever	nts (follow up: mea	n 6 months)										
1 e	randomised trials	not serious	not serious	serious °	very serious <sup>1</sup>	none <sup>a</sup>	19/25 (76.0%)	17/24 (70.8%)	<b>RR 1.073</b> (0.765 to 1.505)	5 more per 100 (from 17 fewer to 36 more)	⊕⊖⊖ VERY LOW	CRITICAL

CI: Confidence interval: OR: Odds ratio: RR: Risk ratio

- a. All studies were described as randomised trials and use of Perchlozone may be associated with patient characteristics. In addition, two publications were described as theses and it is not clear how these were peer reviewed.
- b. Heterogeneity estimated using simulated I squared (39%; p=0.286)
- c. All studies were carried out in the Russian Federation and the majority were carried out in Saint Petersburg, therefore the population under study is restricted to one country
- d. CI does not exclude an appreciable harm and all studies were small in nature (in terms of the population included in the study)
- e. We were only able to extract adverse events data from one trial which is a report of one of the four trials included for the outcome of culture conversion
- f. These data were from one randomised controlled trial with small numbers in both the intervention and control arms (n=25 in the intervention arm, n=24 in the control arm). The confidence interval crosses one

Date: 16 July 2018

Question: PICO 3. In patients with rifampicin-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines are outcomes safely improved with fewer or more than five effective medicines in the intensive phase?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The data were derived from a subset of 8,957 patients from 47 studies included in the main IPD-MA dataset.

			Certainty a	ssessment			Nº of p	patients	Effec	t	Certainty	Importance
N₂ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	five effective medicines in the intensive phase (COMPARATOR)	more or less agents	Relative (95% CI)	Absolute (95% CI)		
Treatment Fa	ilure/Relapse vs. T	reatment Success: 4	Possibly Effective Dru	ugs vs. 5 Possibly Effe	ective Drugs in the In	tensive Phase						
47	observational studies	not serious	not serious	not serious	not serious	none	206/1866 (11.0%)	205/1593 (12.9%)	<b>aOR 1.0</b> (0.7 to 1.3)	0 fewer per 100 (from 4 fewer to 3 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. Tre	atment Success: 4	Possibly Effective D	rugs vs. 5 Possibly Eff	fective Drugs in the In	tensive Phase					•		
47	observational studies	not serious	not serious	not serious	not serious	none	304/1964 (15.5%)	336/1724 (19.5%)	<b>aOR 1.1</b> (0.9 to 1.5)	2 more per 100 (from 2 fewer to 5 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Treatment Fa	ilure/Relapse vs. T	reatment Success: 6	Possibly Effective Dru	ugs vs. 5 Possibly Effe	ective Drugs in the Int	tensive Phase				!		
47	observational studies	not serious	not serious	not serious	not serious	none	206/1866 (11.0%)	87/851 (10.2%)	<b>aOR 1.0</b> (0.6 to 1.7)	0 fewer per 100 (from 4 fewer to 4 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. Tre	atment Success: 6	Possibly Effective D	rugs vs. 5 Possibly Ef	fective Drugs in the In	tensive Phase							
47	observational studies	not serious	not serious	not serious	not serious	none	304/1964 (15.5%)	94/858 (11.0%)	<b>aOR 0.9</b> (0.6 to 1.3)	1 fewer per 100 (from 5 fewer to 2 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL

reatment F	Failure/Relapse vs.	Treatment Success:	2 Highly Effective Dru	gs (INTERVENTION	COLUMN) vs. 1 Highl	y Effective Drugs (COMPARATOR	COLUMN) in the Intensi	ive Phase				
17	observational studies	not serious	not serious	not serious	not serious	none	179/2204 (8.1%)	48/441 (10.9%)	<b>aOR 0.9</b> (0.5 to 1.6)	1 fewer per 100 (from 5 fewer to 4 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. Tr	reatment Success:	2 Highly Effective Dru	ugs (INTERVENTION	COLUMN) vs. 1 High	lly Effective Drugs (CC	DMPARATOR COLUMN) in the Int	rensive Phase	L	l	<u> </u>		· L
47	observational studies	not serious	not serious	not serious	not serious	none	275/2300 (12.0%)	40/433 (9.2%)	<b>aOR 0.60</b> (0.30 to 0.96)	5 fewer per 100 (from 11 fewer to 0 more) <sup>a</sup>	⊕⊕⊖ Low	CRITICAL
Treatment F	ailure/Relapse vs.	Treatment Success:	3 or 4 Highly Effective	Drugs (INTERVENT)	ON COLUMN) vs. 1 h	Highly Effective Drugs (COMPARA	TOR COLUMN) in the In	tensive Phase				1
47	observational studies	not serious	not serious	not serious	not serious	none	179/2204 (8.1%)	14/177 (7.9%)	<b>aOR 0.6</b> (0.2 to 1.8)	3 fewer per 100 (from 11 fewer to 4 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. Tr	reatment Success:	3 or 4 Highly Effective	e Drugs <mark>(INTERVENT</mark>	TON COLUMN) vs. 1	Highly Effective Drug	s (COMPARATOR COLUMN) in the	e Intensive Phase			<b>↓</b>		
47	observational studies	not serious	not serious	not serious	not serious	none	275/2300 (12.0%)	14/177 (7.9%)	<b>aOR 0.40</b> (0.20 to 1.01)	8 fewer per 100 (from 15 fewer to 0 fewer) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Treatment F	Failure/Relapse vs.	Treatment Success:	2 Possibly Effective D	rugs <mark>(INTERVENTIO</mark>	N COLUMN) vs. 3 Pos	L ssibly Effective Drugs (COMPARA	TOR COLUMN) in the Co	ntinuation Phase	L			
47	observational studies	not serious	not serious	not serious	not serious	none	180/1642 (11.0%)	119/719 (16.6%)	<b>aOR 1.3</b> (0.8 to 2.1)	3 more per 100 (from 3 fewer to 9 more) a	⊕⊕⊖⊖ Low	CRITICAL
Death vs. Tr	reatment Success:	2 Possibly Effective [	Drugs <mark>(INTERVENTIO</mark>	N COLUMN) vs. 3 Po	ossibly Effective Drugs	(COMPARATOR COLUMN) in th	e Continuation Phase	<u>.</u>	!	<b>.</b>		
47	observational studies	not serious	not serious	not serious	not serious	none	286/1748 (16.4%)	147/747 (19.7%)	<b>aOR 1.3</b> (0.8 to 2.0)	3 more per 100 (from 4 fewer to 10 more) <sup>a</sup>	⊕⊕©© Low	CRITICAL
Treatment F	Failure/Relapse vs.	Treatment Success:	4 Possibly Effective D	I rugs <mark>(INTERVENTIO</mark> I	N COLUMN) vs. 3 Pos	I ssibly Effective Drugs (COMPARA	TOR COLUMN) in the Co	ntinuation Phase				•
47	observational studies	not serious	not serious	not serious	not serious	none	180/1642 (11.0%)	172/1714 (10.0%)	<b>aOR 1.2</b> (0.9 to 1.5)	1 more per 100 (from 2 fewer to 4 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
		Į	ļ			Ļ	Į	ļ		1		1

Death vs. Tre	eatment Success: 4	4 Possibly Effective D	rugs <mark>(INTERVENTION</mark>	N COLUMN) vs. 3 Pos	ssibly Effective Drugs	(COMPARATOR COLUMN) in the	e Continuation Phase					
47	observational studies	not serious	not serious	not serious	not serious	none	286/1748 (16.4%)	240/1782 (13.5%)	<b>aOR 1.0</b> (0.8 to 1.3)	0 fewer per 100 (from 3 fewer to 4 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Treatment Fa	ailure/Relapse vs. 1	Freatment Success: 2	or More Highly Effect	ive Drugs (INTERVE	NTION COLUMN) vs.	1 Highly Effective Drugs (COMPA	RATOR COLUMN) in the	e Continuation Phase				
47	observational studies	not serious	not serious	not serious	not serious	none	139/2123 (6.5%)	13/253 (5.1%)	<b>aOR 0.5</b> (0.1 to 1.7)	2 fewer per 100 (from 7 fewer to 2 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Death vs. Tre	eatment Success: 2	2 or More Highly Effec	tive Drugs (INTERVE	NTION COLUMN) vs	. 1 Highly Effective D	rugs <mark>(Comparator Column)</mark> i	n the Continuation Phase	2				
47	observational studies	not serious	not serious	not serious	not serious	none	218/2202 (9.9%)	16/256 (6.3%)	<b>aOR 0.8</b> (0.2 to 2.5)	1 fewer per 100 (from 7 fewer to 4 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL

CI: Confidence interval

## Explanations

a. Adjusted risk difference was calculated via fixed effects model.

Date: 16 July 2018

Question: PICO 4. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines are outcomes safely improved with an intensive phase shorter or longer than eight months?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The data subset for this question included records from 3,750 patients for the primary analyses, from 42 observational studies; of whom 2,720 were treated with an individualized MDR-TB regimen and 1,030 were treated with standardized MDR-TB regimens

			Certainty a	ssessment			№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	an intensive phase of 8 months	an intensive phase of less than 8 months	Relative (95% CI)	Absolute (95% CI)		
Treatment Fa	ailure/Relapse vs. 1	reatment Success: 5	-5.99 months on an in	jectable vs. 7-8.49 m	onths on an injectable	2						
42	observational studies	not serious	not serious	not serious	not serious	none	38/623 (6.1%)	145/730 (19.9%)	<b>aOR 0.9</b> (0.4 to 2.3)	0 fewer per 100 (from 3 fewer to 3 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL
Treatment Fa	ailure/Relapse vs. 1	Freatment Success: 6	-6.99 months on an in	jectable vs. 7-8.49 m	onths on an injectable	2						
42	observational studies	not serious	not serious	not serious	not serious	none	38/623 (6.1%)	24/554 (4.3%)	<b>aOR 0.2</b> (0.0 to 1.1)	3 fewer per 100 (from 7 fewer to 0 more) <sup>a</sup>	⊕⊕⊖⊖ Low	CRITICAL

CI: Confidence interval

## **Explanations**

a. Adjusted risk difference comes from fixed effects model.

Date: 16 July 2018

Question: PICO 5. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines are outcomes safely improved with a total duration shorter or longer than twenty months?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The data subset for this question included records from 6,356 patients for the primary analyses, from 51 observational studies; of whom 5,352 were treated with an individualized MDR-TB regimen and 1,004 were treated with standardized MDR-TB regimens

	Certainty assessment						N₂ of patients		Effect		Certainty	Importance
N₂ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	18 to 20 months	total duration more or less than 18 to 20 months	Relative (95% CI)	Absolute (95% CI)		
Treatment Fa	ailure/Relapse vs. T	reatment Success: T	reatment Duration of 2	20-21.99 months vs. 7	Freatment Duration of	17.50-19.99 months						
51	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	31/1376 (2.3%)	87/1407 (6.2%)	<b>aOR 2.1</b> (0.7 to 6.1)	2 more per 100 (from 0 fewer to 3 more) <sup>b</sup>	⊕⊖⊖ VERY LOW	CRITICAL

CI: Confidence interval

## **Explanations**

a. Wide confidence interval

b. Adjusted risk differences calculated from fixed effects model

Date: 16 July 2018

Question: PICO 6. In patients with rifampicin-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines what is the minimum duration of treatment after culture conversion that is more likely to improve outcomes?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The data subset for this question included records from 4,175 patients for the primary analyses, from 39 observational studies; of whom 3 were treated with an individualized MDR-TB regimen and 4,172 were treated with standardized MDR-TB regimens

			Certainty a	ssessment			Nº of p	atients	Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	a treatment duration of 15-17 months after culture conversion	shorter or longer duration	Relative (95% CI)	Absolute (95% CI)		
Treatment Fa	ailure/Relapse vs. 1	reatment Success: T	reatment Duration of 1	12-15 Months Post-Ci	ulture Conversion vs.	Treatment Duration of 15-17 Mon	hs Post-Culture Convers	ion				
39	observational studies	not serious	not serious	not serious	serious <sup>a</sup>	none	9/537 (1.7%)	23/278 (8.3%)	<b>aOR 5.5</b> (0.9 to 33.7)	6 more per 100 (from 2 more to 11 more) <sup>b</sup>	⊕⊖⊖⊖ VERY LOW	CRITICAL
Treatment Fa	ailure/Relapse vs. 1	reatment Success: T	reatment Duration of 1	17-19 Months Post-Ci	ulture Conversion vs.	Treatment Duration of 15-17 Mon	ths Post-Culture Convers	ion				
39	observational studies	not serious	not serious	not serious	not serious	none	9/537 (1.7%)	20/1118 (1.8%)	<b>aOR 1.2</b> (0.4 to 3.7)	0 fewer per 100 (from 2 fewer to 2 more) <sup>b</sup>	⊕⊕CC Low	CRITICAL

CI: Confidence interval

- a. Fixed effects aOR reported due to non-convergence of random-effects model.
- b. Adjusted risk difference calculated using fixed effects.

Date: 30 July 2018

Question: PICO 7. In patients with MDR/RR-TB treated with longer or shorter regimens composed in accordance with WHO guidelines, is monitoring using monthly cultures, in addition to smear microscopy, more likely to detect non response to treatment?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The data were derived from a subset of 3,762 patients treated with longer MDR-TB regimens between 2010 and 2015 in South Africa who had both monthly smear and culture data throughout treatment.

### 1. culture once a month vs smear only once a month

monthly cu	lture	monthly smear			
Sensitivity	0.93 (95% CI: 0.87 to 0.97)	Sensitivity	0.51 (95% CI: 0.41 to 0.60)		
Specificity	0.97 (95% CI: 0.96 to 0.97)	Specificity	0.99 (95% CI: 0.98 to 0.99)		

Prevalences	3.03%	

			F	- actors that ma	ay decrease ce	rtainty of evic	lence		er 1,000 s tested	Test accuracy CoE
Outcome	№ of studies (№ of patients)	Study design				pre-test probability of 3.03%		332		
			Risk of bias	Indirectnes s	Inconsistenc y	Imprecisio n	Publication bias	monthly culture	monthly smear	
True positives (patients with treatment failure)	1 studies 3762 patients	cohort & case-control	not serious	serious <sup>a</sup>	not serious	serious b	none	28 (26 to 29)	15 (13 to 18)	⊕⊕○○ LOW
		type studies						13 more monthly		
False negatives (patients incorrectly								2 (1 to 4)	15 (12 to 17)	
classified as not having treatment failure)								13 fewer monthly		
True negatives (patients without treatment failure)	1 studies 3762 patients	cohort & case-control	not serious	serious <sup>a</sup>	not serious	serious <sup>b</sup>	none	940 (933 to 944)	957 (953 to 961)	⊕⊕⊖⊖ Low
		type studies						17 fewer monthly		
False positives (patients incorrectly								30 (26 to 37)	13 (9 to 17)	
classified as having treatment failure)								17 more i		

- a. The population included had an HIV-positive prevalence of 60.2%.
- b. The data is drawn from a single data set from South Africa.

Date: 30 July 2018

Question: PICO 7. In patients with MDR/RR-TB treated with longer or shorter regimens composed in accordance with WHO guidelines, is monitoring using monthly cultures, in addition to smear microscopy, more likely to detect non response to treatment?

Setting: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries. The data were derived from a subset of 3,762 patients treated with longer MDR-TB regimens between 2010 and 2015 in South Africa who had both monthly smear and culture data throughout treatment.

### 2. culture once a month vs culture every two months

monthly cu	Iture	culture every two months				
Sensitivity	0.93 (95% CI: 0.87 to 0.97)	Sensitivity	0.73 (95% CI: 0.64 to 0.81)			
Specificity	0.97 (95% CI: 0.96 to 0.97)	Specificity	0.98 (95% CI: 0.98 to 0.99)			

Prevalences	3.03%	

			F	-actors that m	ay decrease ce	rtainty of evic	lence		per 1,000 ts tested	
Outcome	№ of studies (№ of patients)	Study design							probability .03%	Test accuracy CoE
			Risk of bias	Indirectnes s	Inconsistenc y	Imprecisio n	Publication bias	monthly culture	culture every two months	
True positives (patients with treatment failure)	1 studies 3762 patients	cohort & case-control	not serious	serious <sup>a</sup>	not serious	serious <sup>b</sup>	none	28 (26 to 29)	22 (19 to 24)	⊕⊕○○ Low
		type studies						6 more T monthly		
False negatives (patients incorrectly								2 (1 to 4)	8 (6 to 11)	
classified as not having treatment failure)								6 fewer F monthly		
True negatives (patients without treatment failure)	1 studies 3762 patients	cohort & case-control	not serious	serious <sup>a</sup>	not serious	serious <sup>b</sup>	none	940 (933 to 944)	954 (950 to 958)	⊕⊕⊖⊖ Low
		type studies						14 fewer monthly		
False positives (patients incorrectly								30 (26 to 37)	16 (12 to 20)	
classified as having treatment failure)								14 more monthly		

- a. The population included had an HIV-positive prevalence of 60.2%.
- b. This data is derived from a single data set from South Africa.

## Annex 9. GRADE evidence to decision tables

## **PICO QUESTION 1**

In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) is a shorter treatment regimen (9-12 months) more or less likely to safely improve outcomes than longer regimens conforming to WHO guidelines?

POPULATION:

The treatment of eligible rifampicin or multidrug-resistant (MDR/RR-TB) patients using a standardized shorter MDR-TB regimen, combining the final results from the STREAM Stage 1 trial and a meta-analysis of pooled individual patient data from observational studies and programmatic cohorts updated in 2018 (2018 IPD SR MA). No data from variants of the shorter regimen in which the injectable agent was replaced by bedaquiline were reported to WHO ahead of the 2018 guideline update.

**INTERVENTION:** 

A standardized shorter treatment regimen (9-12 months)

**COMPARISON:** 

Longer MDR-TB regimens conforming to WHO guidelines

**MAIN OUTCOMES:** 

STREAM Stage 1 trial - (i) Time to culture conversion by week 20; (ii) Favourable outcome; (iii) Died from any cause during treatment or follow-up, all cases; (iiii) Died from any cause during treatment or follow-up, people living with HIV; (iv) Lack of conversion, reversion or relapse; (v) New onset QTcF interval prolongation to 500ms or more on electrocardiogram (automated measurement); (vi) Adverse event of grade 3 to 5 severity (1),(2)

2018 IPD SR MA - analysis for shorter regimens - (i) treatment failure or relapse versus success (cure or treatment completed); (ii) death during treatment vs success; (iii) loss to follow-up vs success, treatment failure or relapse (see Annex 10).

**SETTING:** 

Treatment of adults with MDR/RR-TB using shorter treatment regimens, in low- and middle-income settings, within hospital or ambulatory models of care from trials and observational studies. The studies were conducted in a variety of sites in African and Asian countries (STREAM Stage 1 trial in seven sites in Ethiopia, Mongolia, South Africa and Viet Nam and 2018 IPD SR MA in Bangladesh, Benin, Burkina Faso, Burundi, Cameroon, Côte d'Ivoire, Central African Republic, Democratic Republic of Congo, Eswatini, Kyrgyzstan, Niger, Rwanda, Tajikistan, South Africa and Uzbekistan)

**PERSPECTIVE:** 

The findings of the STREAM trial and the 2018 IPD SR MA are expected to influence the continued validity in its present form of the conditional recommendation made by WHO in 2016 on the basis of an earlier pooled IPD MA of observational studies to use a shorter regimen for MDR/RR-TB (adults and children with pulmonary disease, not previously treated for MDR-TB, not in pregnant women, without resistance to second-line medicines)(3). The role of the shorter MDR-TB regimen in TB programmes is expected to change given the important changes to the composition of longer MDR-TB regimens being recommended in the 2018 guidelines update (see also PICO questions 2 and 3). (for GRADE summary of evidence table see Annex 8).

**BACKGROUND:** 

The interest in reducing the duration of treatment for MDR-TB has motivated a number of initiatives in recent years to treat patients with shorter regimens under programmatic as well as trial conditions(1),(4),(5),(6),(7),(8). When used in carefully selected MDR-TB patients not previously exposed or having strains resistant to second-line medicines, these regimens have been reported to achieve relapse-free cure in over 85% of cases even under programmatic conditions. In 2016, on the basis of data from observational studies of the shorter regimens in different Asian and African countries, WHO recommended a standardized shorter MDR-TB regimen based on the ones under study for eligible patients(3). At that time, the Guideline Development Group assessing the evidence and formulating the recommendation using the GRADE method, proposed a conditional recommendation based on very low certainty in the estimates of effect. By the end of 2017, 62 countries reported having introduced the shorter MDR-TB regimen and about 10,000 patients were reported to have been started on shorter regimens that year alone (9).

In October 2017, the STREAM trial principal investigators presented the preliminary findings of the study during the 48th Union World Conference on Lung Health(4). STREAM Stage 1 was a phase III, multi-centre, international, parallel group, open-label, randomized controlled trial of a standardized MDR-TB treatment regimen lasting 9-11 months versus a longer regimen using a non-inferiority design. The trial enrolled patients between July 2012 and June 2015 (intention to treat (ITT) population = 424 [282 in study arm; 142 control arm]; modified ITT (mITT) population = 369 [245 in study arm; 124 control arm]). Treatment allocation was not blinded to the participants, care givers or data managers. All local and reference laboratory assessments, including microbiological tests involved in the assignment of patient outcome, were conducted blind. When the preliminary data were presented, the findings led to public debate and queries regarding their implications for continued use of the regimen under programmatic conditions, particularly among PLHIV in whom deaths were higher in the study arm than in the control arm. On the basis of the preliminary results WHO issued a position statement, recommending the continued use of the shorter MDR-TB regimen until a full update of the MDR-TB treatment guidelines is completed later in the year(2). The final outcomes of the STREAM trial have been much awaited because they will provide additional information on the efficacy and safety of the shorter regimen and the data are expected to improve the certainty of the estimates (i.e. quality of evidence). In July 2018 the final results of the STREAM trial were made available to WHO. In the analyses of these data, the main observation was that both the shorter and the control regimens obtained a high level of success, even if favourable outcomes were slightly higher in the control regimen (78.8% vs 79.8% in the mITT population). The upper limit of the confidence interval did not reach 10% upon adjustment, thus showing non-inferiority of

A public call for data launched by WHO in February 2018 invited national authorities and technical agencies to submit IPD for both shorter and longer MDR regimen cohorts to inform the 2018 guidelines update (10). As a result of this call, pooled IPD from MDR/RR-TB patients enrolled on standardized shorter regimens between 2005 and 2017 in observational studies or under programmatic conditions in 15 countries were compiled. The main analysis included a maximum of 2,625 records from the shorter regimen studies and 2,717 records from 39 studies of patients on the longer MDR-TB regimens from the separate IPD used to answer PICO questions 2-7 (a description of this IPD is provided in the respective evidence to decision frameworks). No data from variants of the shorter regimen in which the injectable agent was replaced by bedaquiline were reported to WHO while the 2018 guideline update was being prepared.

# **ASSESSMENT**

- 100 = 00111									
Problem Is the problem a p	priority?								
JUDGEMENT	RESEARCH EVIDENCE							ADDITIONAL CONSIDERATIONS	
o No o Probably no o Probably yes • Yes o Varies o Don't know	About 558,000 new cases of rifar cases die(9). Globally only about recent years.  Outcomes of MDR-TB treatment outcome at the end of treatmen and chronicity. Shorter regimens	one fourth of on a global le t. The need fo	newly emergent MDR-T vel are poor with much r better treatment for th	IB patients have been rep loss to follow up and dea nese patients is thus a pri	oorted to start a sec th, and only 55% o ority in an effort to	cond-line TB treatment f cases reported to had a save lives and reduc	nt annually in	MDR-TB is a global challenge and access to treatment often problematic, with regimens typically being long, toxic and expensive.  There is clearly an interest in reducing the duration of treatment, simplifying the administration of the regimen, and providing patients with a safer combination of medicines that can cure the large majority of cases.	
Desirable B How substantial a	Effects  Ire the desirable anticipated effects	?							
JUDGEMENT	RESEARCH EVIDENCE							ADDITIONAL CONSIDERATIONS	
• Trivial	i) STREAM STAGE 1 trial		STREAM Stage 1 trial						
o Small o Moderate o Large o Varies	Outcomes	Relative effect (95% CI)	Anticipated abs	olute effects* (95º	% CI)	Certainty of the evidence (GRADE)	What happens	Based on the results from the STREAM trial, the favourable outcome of the control (longer) regimen is slightly better than the one of the	
o Don't know			Without a standardised 9 month shorter MDR-TB regimen	With a standardised 9 month shorter MDR-TB regimen	Difference			shorter regimen, albeit the effects were not significantly different. These effects do not take into account the potential benefit of the shorter regimen of reducing the treatment duration by half.	
	Time to culture conversion by week 20	<b>HR 1.13</b> (0.91 to	Study population			⊕⊕⊕○ MODERATE <sup>b,c,d</sup>		The final results of the STREAM Stage 1 trial showed that in the mITT and per protocol populations the shorter regimen was not inferior to the leaves are income.	
	assessed with: STREAM Stage 1 trial data (mITT population) № of participants: (1 RCT) <sup>a</sup>	1.40)	0.0% <sup>e</sup>	NaN% (NaN to NaN) <sup>f</sup>	 ( to)			inferior to the longer regimens used in comparison. The difference was marginal but within the limits as defined in the study protocol. A study of a larger sample size could have provided a more definitive estimate.  Nonetheless, the GDG remarked that the	
	Favourable outcome assessed with: STREAM	RR 0.99 (0.88 to	Study population			⊕⊕⊕○ MODERATE <sup>c</sup>		proportion of patients with a favourable treatment outcome was very similar between the intervention and comparator arms and the	
	Stage 1 trial data (mITT population) follow up: mean 132 weeks № of participants: 369	1.01) <sup>h,i</sup>	79.8%	<b>79.0%</b> (70.3 to 80.6)	<b>0.8% fewer</b> (9.6 fewer to 0.8			absolute difference of 0.8% is small and does not imply that the shorter regimen will always perform worse than the longer regimen. The clinical significance of these results needs to be interpreted in the light of a number of factors,	

(1 RCT) <sup>g</sup>				more)	
Favourable outcome assessed with: STREAM	<b>RR 1.02</b> (0.90 to	Study populat	ion		⊕⊕⊕○ MODERATE
Stage 1 trial data (per protocol) follow up: 132 № of participants: 310 (1 RCT) <sup>9</sup>	1.15) <sup>h,j</sup>	(72.7 to 92.8) (8.1 to 12		1.6% more (8.1 fewer to 12.1 more)	MODERATE
Died from any cause during treatment or	<b>HR 1.38</b> (0.64 to	Study populat	ion		⊕⊕⊖⊖ Low <sup>n</sup>
follow-up, all cases assessed with: STREAM Stage 1 trial data (ITT safety population) follow up: mean 132 weeks № of participants: 423 (1 RCT) <sup>k,l</sup>	2.96) <sup>m</sup>	6.4%	<b>8.7%</b> (4.1 to 17.7)	2.3% more (2.2 fewer to 11.4 more)	LOW
Died from any cause during treatment or	<b>HR 2.23</b> (0.76 to	Study populat	ion	⊕⊕○○ LOW <sup>n</sup>	
follow-up, people living with HIV assessed with: STREAM Stage 1 trial data (ITT safety population) follow up: mean 132 weeks Nº of participants: 153 (1 RCT)°	6.60) <sup>m</sup>	8.0%	<b>17.0%</b> (6.1 to 42.3)	9.0% more (1.9 fewer to 34.3 more)	Low
Lack of conversion, reversion or relapse	<b>RR 1.88</b> (0.84 to	Study populat	ion		⊕⊕⊕○ MODERATE
assessed with: STREAM Stage 1 trial data (mITT population) follow up: mean 132 weeks № of participants: 369 (1 RCT) <sup>k,p</sup>	4.21) <sup>h</sup>	5.6%	<b>10.6%</b> (4.7 to 23.8)	<b>5.0% more</b> (0.9 fewer to 18.1 more)	
Lack of conversion,	RR 1.25	Study populat	ion		⊕⊕⊕○

such as the unlikelihood of replicating the rigour maintained in the trial when selecting and following patients up in a programmatic setting; the lack of blinding of the study; the flexibility to modify the treatment in the control arm; no use of bedaquiline, linezolid or clofazimine in the comparator regimens; and the advantage of administering a shorter treatment.

Unadjusted relative risks for lack of conversion, reversion or relapse in patients with baseline resistance to pyrazinamide was 5.95 (95% CL: 0.80, 44.15) and baseline resistance to ethionamide was 3.41 (95% CL: 0.20, 58.32). The 9% excess deaths in persons living with HIV infection in the STREAM trial is an important signal that warrants further investigation in the future.

### 2018 IPD SR MA

The 2018 IPD SR MA showed comparable levels of treatment success in patients treated with the shorter regimen when compared with patients of similar profile treated with longer regimens composed as per WHO recommendations. However, after adjustment, the odds ratio was 2.0 for treatment failure or relapse in the shorter regimen when compared with the longer regimen and 1.2 for death. These effects were largely reproduced in all of the main subgroup analyses done: when longer regimen included Bdq, Lzd or Dlm (OR fail/relapse 9.1; OR death 1.4); PLHIV (2.1; 1.0); PZA-resistance/FQ-susceptible (10.7; 0.3); ETOresistance/FQ-susceptible (3.9; 1.5); extensive disease (1.2; 1.0). The shorter regimen showed better retention than the longer regimen (statistically significant) overall and in all subgroups, presumably a direct consequence of its briefer duration. Decreased loss to follow up may be an important consideration in contexts where this is high.

In conclusion, the shorter regimen has shown effect in both trial and programmatic conditions. It is intended for use in patients who are eligible (e.g. no previous MDR-TB treatment; no strain resistance to fluoroquinolones and second-line injectable agents; no pregnancy).

reversion or relapse assessed with: STREAM Stage 1 trial data (per protocol) follow up: 132 Nº of participants: 310 (1 RCT) <sup>k</sup>	(0.56 to 2.80) <sup>h</sup>	8.4%	<b>10.5%</b> (4.7 to 23.6)	<b>2.1% more</b> (3.7 fewer to 15.2 more)	MODERATE
New onset QTcF interval prolongation to 500ms	<b>RR 2.17</b> (0.91 to	Study population			⊕⊕⊕⊖ MODERATE <sup>r</sup>
or more on electrocardiogram (automated measurement) assessed with: STREAM Stage 1 trial data (ITT safety population) Nº of participants: 423 (1 RCT) <sup>q</sup>	5.14) <sup>h</sup>	4.3%	<b>9.2%</b> (3.9 to 21.9)	<b>5.0% more</b> (0.4 fewer to 17.6 more)	
Adverse event of grade 3 to 5 severity	<b>RR 1.06</b> (0.85 to	Study population			⊕⊕⊕⊜ MODERATE <sup>u</sup>
assessed with: STREAM Stage 1 trial data (ITT safety population) follow up: mean 132 weeks Nº of participants: 423 (1 RCT) <sup>s</sup>	1.32) <sup>h,t</sup>	45.4%	<b>48.1%</b> (38.6 to 59.9)	2.7% more (6.8 fewer to 14.5 more)	

a. As per report to WHO by STREAM Stage I trial investigators on 6 July 2018

- b. The patient group is expected to be similar to the one who would receive the regimen but there was insufficient information about risk groups at the time of the review in January 2018 (e.g. frequency of cavitary disease on radiography, disease severity, exposure to second-line treatment, ARV in PLHIV). No trial data in children. Overall the trial eligibility criteria are similar to the ones recommended by WHO in the guidelines and there was no restriction in CD4 count among the PLHIV in the trial (one third of study participants). The main concern would be to extrapolate to patients presumed to have strains susceptible to fluoroquinolones and 2nd line injectable agents but in whom resistance has not been reliably excluded by DST.
- c. Single trial, number of patients in each arm is relatively small with wide confidence limits.
- d. The risk of bias was not downgraded for lack of blinding of patients and clinicians because this was an inherent study design and measures were taken to mitigate the blinding; cross-over from shorter to a longer regimen was also very low. No information provided about selection bias upstream of randomization
- e. Kaplan-Meier estimate of culture conversion at Week 20: 99.2% (95% CL: 97.3%, 99.8%)
- f. Kaplan-Meier estimate of culture conversion at Week 20: 99.2% (95% CL: 95.8%, 99.9%)
- g. Favourable outcome is defined as culture negative at 132 weeks and at the previous occasion that the patient was seen, unless the patient outcome had already been classified as unfavourable (see below). The mapping of trial endpoints to the WHO TB outcomes definitions (2005 or 2013) was not direct (e.g. some cases normally classified as Cured could be reclassified as unfavourable in the trial should death or relapse happen during the

In settings like eastern European/central Asian where resistance patterns are more advanced and where performance of DST to some of the medicines is challenging, the regimen may be expected to be less effective. The excess mortality in PLHIV in the trial is an important signal that warrants further investigation in future.

The GDG acknowledged that the efficacy of shorter regimens may differ according to the sub-group being treated, and for some sub-groups there is limited evidence, namely

- Resistance to agents in the shorter regimen, other than isoniazid resistance
- People living with HIV infection
- Exclusive extra-pulmonary disease, particularly serious forms such as CNS and disseminated TB and extrapulmonary disease in PLHIV
- Children
- Pregnant women

follow-up period).

- h. Unadjusted value
- i. The difference in proportions (1.1% (-7.7%, 9.8%)) becomes smaller after adjustment for HIV status (1.0% (-7.5%, 9.5%)) but is still within the limits of non-inferiority.
- j. The difference in proportions (-1.2% (-11.1%, 8.6%)) becomes smaller after adjustment for HIV status (-0.7% (-10.5%, 9.1%)).
- k. This outcome was a subset from among the "unfavourable outcomes" of the trial. "Unfavourable outcomes" were defined as not satisfying the favourable outcome criteria (see above) because of (i) start of 2 or more additional medicines (including a change of regimen); or (ii) treatment extended beyond the permitted duration; or (iii) death at any point up to 132 weeks post-randomization; or (iv) positive culture result at 132 weeks post-randomization or when last seen; or (v) not seen at 76 weeks or later
- I. In the mITT population the difference in mortality between shorter and longer regimen arms was also not statistically significant (9.5% vs 6.9% respectively; unadjusted RR 1.37 (0.66 to 2.86))
- m. Time to death
- n. Imprecision downgraded by two levels because the span of the confidence interval implies important uncertainty on potential risk of death or survival
- o. In patients without HIV the unadjusted RR for death was 0.610 (95%CI: 0.191-1.946). However, a subgroup analysis by HIV status was not pre-specified in the STREAM trial protocol.
- p. Unadjusted RR for lack of conversion, reversion or relapse in patients with baseline resistance to pyrazinamide was 5.95 (95% CL: 0.80, 44.15) and baseline resistance to ethionamide was 3.41 (95% CL: 0.20, 58.32)
- q. Moxifloxacin was given at 400mg/day in the longer regimen arm regardless of body weight; patients on the shorter regimen received either 400mg (<33kg body weight), 600mg (33-50kg) or 800mg (>50kg). A QTcF of 500ms+ was more frequent in shorter regimen than in longer regimen patients in both those weighing 33-50kg (9% vs 3%) and >50kg (10% vs 5%) (unadjusted RR 2.17 (95%CL: 0.92, 5.16)).
- r. The availability of electrocardiography in a typical centre treating MDR-TB patients may vary.
- s. Grade of severity was defined by the criteria of the Division of AIDS (DAIDS). Table for grading the severity of adult and pediatric adverse events. Version 1.0, Dec 2004; Clarification August 2009. http://rsc.tech-res.com/docs/default-source/safety/table\_for\_grading\_severity\_of\_adult\_pediatric\_adverse\_events.pdf. Hearing loss was not assessed using audiometry in all centres.
- t. The difference (-2.8% (-12.9%, 7.3%)) hardly changed after adjustment for HIV status (-2.8% (-12.9%, 7.2%))
- u. The expert panel did not consider the variable use of audiometry to assess hearing loss by sites to be a reason to downgrade inconsistency for this outcome.

# ii) 2018 IPD SR MA

# A shorter treatment regimen (9-12 months) compared to longer regimens conforming to WHO guidelines for patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB)

Outcomes	With longer regimens conforming to WHO guidelines	With a shorter treatment regimen (9-12 months)	Difference	Relative effect (95% CI)
Fail/release ve Cuescas Charles ve Langua regimens	C non 100	11 per 100	5 more per 100	OR 2.00
Fail/relapse vs Success: Shorter vs Longer regimens	6 per 100	(6 to 20)	(0 fewer to 14 more)	(0.96 to 4.00)
Death vs Success: Shorter vs Longer regimens	13 per 100	15 per 100	2 more per 100	OR 1.20
Death vs Success. Shorter vs Longer regimens	13 per 100	(12 to 19)	(1 fewer to 6 more)	(0.95 to 1.60)
Last va Sussess Epil/relance, Charter va Langer regimens	21 per 100	5 per 100	16 fewer per 100	OR 0.2
Lost vs Success, Fail/relapse: Shorter vs Longer regimens	21 per 100	(5 to 8)	(16 fewer to 14 fewer)	(0.2 to 0.3)
Fail/relapse vs Success amongst patients with confirmed FQN-susceptible TB: Shorter versus Longer regimens that	4 per 100	28 per 100	24 more per 100	OR 9.1
include Bdq, Lzd, or Delamanid	4 per 100	(3 to 85)	(2 fewer to 81 more)	(0.6 to 135.4)
Death vs Success amongst patients with confirmed FQN-		13 per 100	3 more per 100	OR 1.4
susceptible TB: Shorter versus Longer regimens that include Bdq, Lzd, or Delamanid	10 per 100	(7 to 21)	(3 fewer to 11 more)	(0.7 to 2.5)
Fail/relapse vs Success in People Living With HIV (PLWH)	7 por 100	14 per 100	7 more per 100	OR 2.1
rail/Telapse vs Success iii reople Livilig With Tilv (rLWII)	7 per 100 : 20 per 100 : 2	(5 to 38)	(3 fewer to 31 more)	(0.6 to 7.7)
Death vs Success in PLWH	20 per 100	20 per 100	0 fewer per 100	OR 1.0
Death vs Success in Lewin	20 per 100	(13 to 28)	(7 fewer to 8 more)	(0.6 to 1.6)
Fail/relapse vs Success PZA-Resistant, FQ-S	3 por 100	26 per 100	23 more per 100	OR 10.7
rany relapse vs Success FZA-Nesistant, 1 Q-3	3 per 100	(6 to 68)	(2 more to 65 more)	(1.8 to 64.5)
Death vs Success PZA-Resistant, FQ-S	9 por 100	3 per 100	6 fewer per 100	OR 0.3
Death vs Success FZA-Resistant, TQ-3	20 per 100 ( 2 per 100 ( 3 per 100 ( 9 per 100 ( 1	(1 to 12)	(8 fewer to 3 more)	(0.1 to 1.4)
Fail/relapse vs Success ETO/PTO-Resistance, FQ-S	2 nor 100	10 per 100	7 more per 100	OR 3.9
rail/Telapse vs Success ETO/PTO-Resistance, PQ-S	3 per 100	(3 to 29)	(0 fewer to 27 more)	(1.0 to 15.1)
Donth vs Success ETO/DTO Posistance EO S	6 per 100	9 per 100	3 more per 100	OR 1.5
Death vs Success ETO/PTO-Resistance, FQ-S	o hei 100	(2 to 31)	(4 fewer to 26 more)	(0.3 to 7.4)
Fail/relapse vs Success Extensive	6 per 100	7 per 100	1 more per 100	OR 1.2
i aii/Teiapse vs Success Exterisive	0 per 100	(4 to 15)	(2 fewer to 9 more)	(0.6 to 2.6)
Death vs Success Extensive	12 per 100	13 per 100	0 fewer per 100	OR 1.0
Death vs Success extensive	13 per 100	(11 to 16)	(2 fewer to 3 more)	(0.8 to 1.3)

# **Undesirable Effects**

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Large o Moderate o Small o Trivial  ● Varies o Don't know		The harms are expected to vary by patient subgroup and by the type of adverse event (AE) (i.e. higher frequency of QT-interval prolongation and hepatotoxicity in the STREAM trial regimen).  The shorter regimen does not include a number of second-line agents that are most often associated with serious or distressing AEs (such as cycloserine, PAS, linezolid), and which can thus be reserved for use in a longer regimen should the patient not respond to the shorter regimen or sustain a relapse. However they do contain kanamycin as an integral part of the regimen, given for 4 months or more, and this is has been reported to cause hearing loss even when used in the shorter regimen (11).  The higher level of QT-interval prolongation observed in the patients on the shorter regimen in the STREAM Stage 1 trial may be associated with use of higher dose moxifloxacin (600mg/day in patients 33-50kg and 800mg/day in >50kg).  A higher risk of death was observed among PLHIV in the STREAM trial (9% excess), however these findings were not statistically significant. In the IPD-MA there was no difference in deaths (0% excess) among PLHIV when comparing the shorter to longer regimens. In those who were HIV negative, 1% fewer deaths were reported among persons receiving a shorter regimen, although this was not a statistically significant finding.  The GDG noted that the undesirable effects were moderate for PLHIV but otherwise small for the general population, and therefore concluded that the undesirable effects may vary, depending on the group being treated.

# Certainty of evidence What is the overall certainty of the evidence of effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low  ● Low  o Moderate  o High  o No included studies	Results provided in the GRADE tables are from two sources: (i) the STREAM Stage 1 trial, a single phase III, multi-centre, parallel group, open-label, randomized controlled trial in seven sites in 4 countries, and (ii) an IPD MA which included 2,625 patients from 15 countries, with patients enrolled between 2005 and 2017.  The certainty of the evidence was rated "low" overall but was "moderate" for most of the effects in the STREAM trial (except for death, graded as "low") and "low" or "very low" for the estimates of the IPD MA. The certainty of the evidence has thus improved in comparison to the earlier observational study IPD MA on which the 2016 recommendation was based, which was graded "very low certainty"	The STREAM trial could not be blinded for the patients, care givers or data managers. All local and reference laboratory assessments, including microbiological tests done to assign patient outcome, were conducted blind. While measurement of endpoints was rigorous in most cases, the assessment of hearing loss was deemed unsatisfactory by the expert panel. The data included in the IPD-MA are observational data from programmatic settings.  The GDG considered that the excess deaths among PLHIV observed in the STREAM trial may be an important signal. The risk of dying among PLHIV in the intervention group (17.5% of the safety population) was not statistically significantly different from the control arm (8%). Available information could not confirm or exclude that these deaths could be directly or indirectly attributable to a failure to resolve MDR-TB or to recrudescent disease, which could be averted had treatment been prolonged beyond 9 months. Moreover, the STREAM trial was not powered to detect subgroup effects. However, the multicentre observational study in Africa published in January 2018 also reported a difference in deaths between PLHIV and HIV negative patients (19.0% vs. 5.0% respectively, P<0.001) (8). In this study 89% of PLHIV received anti-retroviral therapy (ART) and the proportion of patients who died did not differ between those who did and did not receive ART (18.6% vs. 19%). In addition, the IPD-MA did not show a difference in the risk of death among PLHIV and in uninfected people.

W	$\sim$ 1	u	

Is there important uncertainty about or variability in how much people value the main outcomes?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Important uncertainty or variability o Possibly important uncertainty or variability o Probably no important uncertainty or variability • No important uncertainty or variability	No research evidence was identified.	The GDG judged that there was no important uncertainty or variability in values attached to the main outcomes (success, failure, death and adverse events). The briefer duration of the shorter regimen was considered an important consideration for the patients and services and increased the likelihood of patient adherence to treatment till the end.

# **Balance of effects**

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Favors the comparison o Probably favors the comparison • Does not favor either the intervention or the comparison o Probably favors the intervention o Favors the intervention o Varies o Don't know		While there was an imbalance in deaths between HIV infected and uninfected study subjects in the STREAM trial there was no clear benefit or harm related to the study regimen that could be plausibly linked or explained by confounding. The IPD MA also supported this finding and indicated also that loss to follow up was reduced significantly. The main disadvantage of the shorter regimen is the need to have an injectable agent in the first 4 to 6 months, in contrast to new oral-only longer regimens that however need to be given for twice as long. The shorter regimen is also standardised and therefore cannot be modified if there is resistance to medicines like pyrazinamide and ethionamide that have been associated with worse outcomes. The GDG concluded that in eligible patients the effects did not differ substantially between the study and control arms to be of clinical and public health relevance (with the exception of loss to follow-up). Therefore, the GDG felt that the balance of effects does not favour either the shorter regimen or the longer regimen.

# **Resources required**

How large are the resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Large costs o Moderate costs o Negligible costs and savings o Moderate savings o Large savings  • Varies o Don't know	Preliminary health economic analysis from the STREAM trial indicate that the ninemonth regimen reduced the cost of treatment for the health system, compared to the 20-month regimen. The cost reduction to the health service per patient was USD 2,879 in Ethiopia - a reduction of 34% - and USD 4,916 in South Africa - a reduction of 46% (factoring in the ECG monitoring in the study arm). Patients on the study regimen also had about 15 outpatient visits less than the controls. The ninemonth regimen also reduced direct costs to the patient (transport and food) as it required fewer visits to health facilities. In Ethiopia this reduced costs by around USD 18 per patient. Patients also had reduced expenditure on supplementary food in the nine-month arm, particularly in weeks 12 to 84 of treatment. This reduction in costs was around USD 140 per patient in Ethiopia. The shorter regimen also allowed patients to return to work sooner.	Although the data on costs collected during the STREAM trial were sourced from the national TB programmes, it was felt that the costs were reliable enough to be able to answer this question. The cost of the medicines needed for one treatment is around USD 700 when purchased from the Global Drug Facility (GDF)(12). The savings were thought to vary, in the range of moderate to large. It is expected that any initial costs for training and strengthening of infrastructure to implement the shorter regimen would be discounted over time.

# Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low	The STREAM trial includes amongst its secondary outcomes (i) costs to the health	Even though this was not formally assessed, detailed information on the cost of medicines in different
o Low	system related to delivering the regimen and conducting follow-up tests, and (ii)	regimens were available from GDF, a major supplier of TB medicines globally. The GDG judged that
Moderate	household costs. The data included in the IPD-MA were not linked to costs or any cost	there was moderate certainty that use of the shorter regimen would reduce resource use due to a
o High	effectiveness analyses.	lowered cost of medicines and patient monitoring/visits.
O No included studies		

# **Cost effectiveness**

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

O Favors the comparison O Probably favors the comparison O Does not favor either  A cost-effectiveness analysis from the STREAM trial is planned but the results were not available to the GDG at the time of this review.	JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
the intervention or the comparison O Probably favors the intervention O Favors the intervention O Varies  No included studies	o Probably favors the comparison o Does not favor either the intervention or the comparison o Probably favors the intervention o Favors the intervention o Varies		

_		
Ŀα	Шť\	1
느띡	шь	ì

What would be the impact on health equity?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
O Reduced O Probably reduced O Probably no impact O Probably increased O Increased O Varies  Don't know	There were no included studies on equity for the GDG to consider.	Although the cost effectiveness analysis from the STREAM trial is pending, reliable data are available on the costs of the shorter TB regimen based on programmatic implementation. The cost of the medicines required for the average adult patient equate to roughly one third of the cost of a typical longer regimen. It is thus expected that both expenses related to the medicines and implementation would not be higher than for longer regimens, implying that within a finite budget additional resources could be released for the treatment of more patients. The intervention could thus favour vulnerable populations. The fact that the regimen requires an injectable and ECG monitoring and active TB drug safety monitoring and management (aDSM) may demand specialised care and monitoring and may thus drive the balance away from equity.  Modelling work based on experience of the use of the shorter regimen in a high MDR-TB burden setting shows a potential to reduce transmission of resistant strains and a positive epidemiological impact, beyond patients who are eligible(13). Greater gains may be possible in contexts where the regimen is more widely applicable and available.

Acceptability
Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o No o Probably no o Probably yes • Yes o Varies o Don't know	treatment and support experiences (frequency of health facility visits, adverse events) and (ii) health worker experience of delivering treatment and support. These data would help appreciate better the acceptability of regimen to all stakeholders but	The shorter MDR-TB regimens have been successfully implemented in a number of settings in Africa and Asia in recent years through the efforts of a number of technical agencies and national programmes. The standardized nature of the shorter regimen facilitates implementation. Clinicians and patients have found the intervention to be acceptable. In the STREAM trial only about 5% of patients in the study arm required a prolongation of treatment (Twelve GDG members voted for Yes; 4 for probably yes; 5 abstained and 2 were absent)
		, , , , , , , , , , , , , , , , , , , ,

# **Feasibility**

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o No o Probably no ● Probably yes o Yes o Varies o Don't know	The GDG concluded that the shorter regimen was likely feasible to implement as it has already been implemented in a number of countries globally. Other feasibility considerations include the availability of ECG and DST.	The expert panel noted that ECG and DST were probably the most important determinants of feasibility. By the end of 2017, 62 countries reported having introduced the shorter MDR-TB regimen and about 10,000 patients were reported to have been started on shorter regimens that year alone (9). The intervention has been rolled out even when resources were constrained and has been supported by major donors such as the Global Fund(14).  Two staples of the shorter MDR-TB regimen - clofazimine and single-dose isoniazid - may be difficult to procure in some countries. Clofazimine, which is used off-label in MDR-TB regimens (given that its marketing authorization is for leprosy), is problematic in Latin America and elsewhere. No quality-assured source of gatifloxacin - a cheap fluoroquinolone which was a cornerstone of the shorter MDR-TB regimen until relatively recently - is available today given a global shortage in manufacture following a reported risk of associated dysglycaemia. Its replacement by moxifloxacin was originally shackled by a

	much higher price but this has since dropped as the patent on this medicine expired a few years ago.
	Given the lack of scoring on the 400mg tablet, it is currently not easy to administer the 600mg dose of moxifloxacin for patients in the 33-50kg weight band when using the higher dose option as in the STREAM trial. Further, social support was provided in different forms in the STREAM trial sites (e.g. Ethiopia, South Africa), which may be feasible in some settings but not elsewhere.

# **SUMMARY OF JUDGEMENTS**

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	Yes		Varies	Don't know
DESIRABLE EFFECTS	Trivial	Small	Moderate	Large		Varies	Don't know
UNDESIRABLE EFFECTS	Large	Moderate	Small	Trivial		Varies	Don't know
CERTAINTY OF EVIDENCE	Very low	Low	Moderate	High			No included studies
VALUES	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability			
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	Don't know
RESOURCES REQUIRED	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	Low	Moderate	High			No included studies
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	No included studies
EQUITY	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know
ACCEPTABILITY	No	Probably no	Probably yes	Yes		Varies	Don't know
FEASIBILITY	No	Probably no	Probably yes	Yes		Varies	Don't know

# **TYPE OF RECOMMENDATION**

Strong recommendation against the	Conditional recommendation against the	Conditional recommendation for either the	Conditional recommendation for the	Strong recommendation for the
intervention	intervention	intervention or the comparison	intervention	intervention

## **CONCLUSIONS**

## Recommendation

In MDR/RR-TB patients who have not been previously treated for more than one month with second-line medicines used in the shorter MDR regimen or in whom resistance to fluoroquinolones and second-line injectable agents has been excluded, the shorter MDR-TB regimen may be used as an alternative to a longer MDR-TB regimen (conditional recommendation; low certainty in the estimates of effect).

## **Justification**

### Overall justification

The analyses performed for the 2018 guidelines showed that patients who received the shorter MDR-TB treatment regimens had a similar likelihood of treatment success when compared to individuals who received longer regimens (1% lower favourable outcomes in the STREAM trial and 2% higher treatment failure or relapse in the IPD-MA; both differences not statistically significant). In both the STREAM trial and the 2018 IPD SR MA, the patients with additional resistance to pyrazinamide or ethionamide/prothionamide experienced higher risk of non-response (treatment failure, relapse, reversion or lack of conversion) these results were not statistically significant and had wide confidence limits. When deciding on the strength and direction of the recommendation, 13 GDG members voted conditional for the shorter regimen, 4 voted conditional against the shorter regimen, 3 members abstained and 3 were not present.

### **Detailed justification**

### Problem

TB is the single most important infectious cause of death and MDR-TB increases the likelihood of dying and other poor outcomes in TB patients. Proper treatment of TB and MDR-TB is important to help reduce the global burden of disease and death. Simpler and shorter regimens for MDR-TB could help improve access and effectiveness of treatment.

#### Desirable Effects

Many MDR-TB patients worldwide would be eligible for the shorter MDR-TB regimen. In these patients this regimen offers the prospect of halving the time to completion of treatment when compared with the other treatment options. Favourable outcomes in eligible individuals are comparable to the ones in the longer regimen (close to 80%). If second-line line probe assay or DST can be applied more universally to MDR/RR-TB patients then the balance of acceptability shifts in favour of the intervention.

### Undesirable Effects

The GDG reviewed the individual information for all deaths to assess the excess mortality in PLHIV (based on the STREAM trial data). The panel agreed that this is an important signal nonetheless but not sufficient grounds for a negative recommendation on use of regimen in PLHIV. The IPD-MA did not show a difference in the risk of death among PLHIV (0% excess) and in people not infected with HIV (1% reduction; not statistically significant). Higher frequency of QT interval prolongation observed in the study arm of the STREAM trial may have been influenced by the higher dose of moxifloxacin used in this study (and possibly the additive effect of clofazimine that also prolongs the QT interval). It is worthy of further investigation, even if no cases of dysrhythmias or cardiac deaths were attributed to the regimen. Monitoring for hearing loss was deemed insufficient in the STREAM trial. The inconvenience of daily, painful intramuscular injections for the first 4-6 months is an important consideration in patients on the shorter regimen. The risk of relapse and acquisition of additional resistance are probably low in patients without drug resistance additional to MDR-TB.

### Certainty of evidence

The overall certainty in the estimates of effect varied from HIGH to VERY LOW, but was commonly in the MODERATE to LOW range for most estimates of effect. When the certainty in the estimates of effect was downgraded it was usually done for imprecision, given that for some outcomes the number of observations was small and 95% confidence intervals were wide.

### Values

The GDG judged that many eligible patients, once informed of the options, would prefer the shorter regimen over the longer one. However, the availability of oral-only regimens inclusive of more effective agents like bedaquiline and linezolid may tip the balance in their favour even if they require twice as much time to complete. A number of recent publications involving patients and carers have advocated strongly against the

continued use of injectable agents in MDR-TB regimens(15),(16). Health care managers and patients would also appreciate a regimen which requires skilled staff to administer repeatedly for several months.

Balance of effects

The GDG felt that the overall benefits of using the shorter regimen would be likely to offset the downsides of using it.

Resources required

The GDG judged that resources are likely to be reduced due to a lowered cost of medicines and patient monitoring.

Certainty of evidence of required resources

Even though this was not formally assessed, detailed information on the cost of medicines in different regimens were available from GDF, a major supplier of TB medicines globally. The GDG judged that there was moderate certainty that use of the shorter regimen would reduce resource use due to a lowered cost of medicines and patient monitoring/visits.

Cost effectiveness

The results of the cost-effectiveness analysis of the STREAM Stage 1 trial were not available at the time of the review.

Equity

This was not formally assessed. However, the GDG felt that availability of the shorter regimen might increase health equity by providing wider access and releasing more resources that can be channelled back into patient care and by allowing patients to return to work and resume other responsibilities earlier.

Acceptability

This was not formally assessed. However, the shorter MDR-TB regimen has been found to be acceptable in many settings.

Feasibility

The shorter MDR-TB regimen has been successfully implemented in many settings, indicating that it is feasible. The availability of electrocardiography may limit implementation and decentralization of the regimen.

## **Subgroup considerations**

When WHO first issued its recommendations on the shorter MDR-TB regimen in 2016, they were accompanied by inclusion criteria (3). Previous treatment with second-line drugs for more than one month, resistance to medicines in the regimen, extrapulmonary disease and pregnancy were exclusion criteria. The recommendation was made subject to patients having been tested for *in vitro* resistance to at least fluoroquinolones and the injectable agent used in the regimen before starting treatment. In the absence of reliable testing, patients who were highly unlikely to be infected with resistant strains based on clinical or recent representative surveillance data were also eligible for the shorter MDR-TB regimen.

In the evidence reviewed for the 2018 guidelines, treatment outcomes in patients with laboratory confirmed resistance to pyrazinamide and ethionamide/prothionamide were poorer than in those without additional resistance. The 2018 recommendation thus reinforces the importance of excluding resistance to fluoroquinolones and second-line injectable agents before the shorter MDR-TB regimen is considered. Other testing, such as DST to pyrazinamide and genotyping studies of isoniazid resistance - are also considered important and should be performed if possible.

Decisions to start newly diagnosed patients on the standardized shorter MDR-TB regimen should be made according to patient preference and clinical judgement, for patients who do not have any of the following conditions:

- 1. Resistance to or suspected ineffectiveness to a medicine in the shorter MDR-TB regimen (except isoniazid resistance);
- 2. Exposure to one or more second line medicines in the regimen for >1 month (unless susceptibility to these second line medicines is confirmed);
- 3. Intolerance to any medicine in the shorter MDR-TB regimen or risk of toxicity from a medicine in the shorter regimen (e.g. drug-drug interactions);
- 4. Pregnancy;

- 5. Disseminated, meningeal or central nervous system TB;
- 6. Any extrapulmonary disease in HIV patients.

Considerations for specific sub-groups are provided below, based on the review of the current evidence by the GDG

People living with HIV: One third of the STREAM trial participants were HIV positive without CD4 restriction. The majority of deaths among PLHIV occurred in two sites in South Africa (19/24 deaths among 151 total participants). The reason for the excess deaths observed in the study arm among PLHIV remains unclear but may be a clinically-relevant signal. A detailed assessment by the expert panel of causality of death in the 33 study subjects (9 of whom were HIV negative) who died during treatment or on follow-up did not reveal any signals that the shorter regimen was associated with harms in PLHIV or an interaction (e.g. excessive pill burden, poor adherence, drug-drug interactions with anti-retroviral medications). Nonetheless it is recommended that the regimen may be used in PLHIV on the condition of careful monitoring of anti-retroviral (ARV) effectiveness. As in any other PLHIV, patients receiving the shorter regimen who also have HIV infection will need prophylactic medication for opportunistic infections, support for TB and ARV medication adherence and close monitoring of the biomarkers of immune status. In the IPD-MA the likelihood of treatment failure and death were similar in PLHIV and the uninfected.

**RR-TB without MDR-TB:** Only 5.8% of STREAM trial participants in the study arm were isoniazid susceptible. All patients – children or adult - with rifampicin-resistant TB in whom isoniazid resistance is not confirmed may be treated with the shorter MDR-TB treatment regimen.

Resistance additional to isoniazid and rifampicin: The STREAM trial underlines the effectiveness of the regimen in patients without resistance to fluoroquinolones and second line injectable agents. STREAM trial data demonstrated a higher unadjusted relative risk of culture reversion, relapse or lack of culture conversion for patients with baseline resistance to pyrazinamide and ethionamide (albeit not statistically significant and with wide confidence limits). The IPD-MA also showed a higher risk of treatment failure and relapse in patients with resistance to pyrazinamide and ethionamide/ prothionamide demonstrated among patients with resistance, when compared to those who were susceptible. In patients having strains with laboratory-confirmed resistance to components in the shorter MDR-TB regimen, or solid grounds to believe that they are ineffective (e.g. contact with a patient with documented resistance), the shorter regimen should not be used. In the absence of reliable test results for regimen components in an individual patient, representative data on the background prevalence of resistance may help decide if the shorter regimen may be used or not. It is also recommended that in areas with a high prevalence of pyrazinamide or ethionamide resistance that alternative regimens be used. Such uncertainty is one of the reasons that the recommendation for the shorter regimen remains conditional.

Children: Children were excluded from the STREAM trial. However there were 78 children and adolescents who received the shorter regimen in the 2018 IPD SR MA. The effect of the shorter regimen on treatment outcomes for children and adolescents has been difficult to determine due to the small numbers for each outcome. However, while there is no plausible biological reason to believe that these regimens are less effective or less tolerable in children than in adults, it is acknowledged that additional data on its use in children would be useful. The avoidance of an injectable-containing regimen is particularly desirable in children, especially those very young and without severe disease. Given the profound impact that hearing loss can have on the acquisition of language and the ability to learn at school, use of injectable agents in children has to be accompanied with regular audiometry. It is recommended that children with pulmonary MDR/RR-TB be given the same consideration for treatment with a shorter MDR-TB treatment regimen as adults.

**Pregnant women:** Pregnancy was an exclusion criterion for the STREAM trial. Two of the components of the shorter MDR-TB regimens – the injectable agent and ethionamide (or prothionamide) – are usually contraindicated in pregnancy. Withholding these medicines from the shorter MDR-TB treatment regimen could seriously compromise its effectiveness. In the case of pregnant females it is therefore recommended that an individualised, longer regimen be used which can allow the selection of four or more effective medicines with lower teratogenic risk.

Exclusive extra-pulmonary disease: The findings from STREAM trial were limited to patients who had pulmonary localisation and they cannot be extrapolated directly to all different forms of extrapulmonary disease. It is proposed that the shorter regimen be avoided in patients with disseminated TB or TB of the central nervous system, as well as in all PLHIV who have extra-pulmonary disease.

**Persons with diabetes mellitus:** There are no data on the use of the shorter regimen among people with diabetes mellitus. It is recommended that patients with diabetes be given the same consideration for treatment with a shorter MDR-TB treatment as for all other patients.

## Implementation considerations

The shorter MDR-TB regimen has become well known in the TB community. Attempts to gradually reduce the duration of the regimen in Bangladesh started two decades ago and stabilised latterly to a 9 month regimen with 7 agents in a 4-month intensive phase and 4 agents in a 5-month continuation phase(5). This regimen, with some variations, was subsequently adopted in other low-resource settings, mostly in Africa, but also in high MDR-TB settings (e.g. Kyrgyzstan, Tajikistan, Uzbekistan). The same regimen - 4-6Km-Mfx-Cfz-Eto-Z-E-Hh / 5Mfx-Cfz-Z-E - was tested out in the STREAM Stage 1 trial, which enrolled patients between 2012 and 2015. In 2016, WHO recommended the use of the shorter MDR-TB regimen subject to specific inclusion/exclusion criteria; since then several countries have introduced the regimen. Given its largely standardised composition and duration the regimen has been relatively easy to implement.

In order to reproduce the high cure rates achieved in the STREAM trial all efforts need to be made to prevent the acquisition of additional resistance, through careful selection of patients to be enrolled, and effective patient support to enable full adherence to treatment. The STREAM trial investigators observed that tolerance to poor adherence was lower in the shorter regimen than in the control arm. It is recommended that patients be tested for susceptibility or resistance to fluoroquinolones and to the second-line injectable agent used in the regimen before being started on a shorter MDR-TB regimen and patients with strains resistant to any of the two groups of medicines transferred to treatment with a conventional regimen. If testing for susceptibility or resistance to other medicines used in the regimen is available (i.e. for pyrazinamide or ethionamide, for instance) it is highly recommended that this also be carried out at baseline.

The availability of reliable and rapid tests is valuable to decide within a few days which patients would be eligible for shorter MDR-TB regimens - or what modifications to longer MDR-TB regimens would be necessary based on resistance detected. In patients with confirmed MDR/RR-TB, the MTBDRsl assay may be used as the initial test, over culture and phenotypic DST, to detect resistance to fluoroquinolones and to the second-line injectable drugs (conditional recommendations; certainty of evidence for direct testing of sputum from low to moderate (17). This applies to testing in both children and adults. While resistance-conferring mutations to fluoroquinolones detected by the MTBDRsl assay are highly correlated with phenotypic resistance to ofloxacin and levofloxacin, the correlation with moxifloxacin is less clear and the inclusion of moxifloxacin in a MDR-TB regimen is best guided by phenotypic DST results (likewise for gatifloxacin, should a quality assured preparation become available in future). In settings in which laboratory capacity for DST to fluoroquinolones and injectable agents is not yet available, the clinician and the TB programme manager would need to decide on the basis of the likelihood of resistance to these medicines, informed by the patient's clinical history and recent representative surveillance data. As rapid DST is not available for all medicines used in the shorter regimen (for example, pyrazinamide, which relies on phenotypic testing) then the short regimen may be started while waiting for these results, and if needed, the patient can be transferred to a longer regimen if additional resistance is detected.

The evidence for the effectiveness and safety of the shorter MDR-TB regimen now derives from both trial sites and observational studies where this treatment was administered under fairly standardised conditions with relatively little variation in the content and duration. The recommendation on the use of the shorter MDR-TB regimen is made under the premise that it is implemented as per the composition and duration used in the studies. Replacement of medicines and prolongation/shortening of the duration would only be permissible within the parameters applied in these studies (e.g. gatifloxacin replaced by moxifloxacin; prothionamide replaced by ethionamide; intensive phase prolonged up to 6 months in case of no sputum conversion). No data from variants of the shorter regimen in which the injectable agent was replaced by bedaquiline were reported to WHO while the 2018 guideline update was in process. Programmes wanting to implement important variations to the shorter regimen (e.g. replacement of the injectable agent with bedaquiline or linezolid) should therefore only do so under operational research conditions.

Given the global shortage in the supply of quality-assured gatifloxacin in recent years, the STREAM trial and the observational studies have had to substitute this agent with high-dose moxifloxacin. This has led to an increase in the overall price of the regimen, with moxifloxacin typically accounting for about one half of overall drug costs even if its cost has recently come down as a result of the availability of more generic preparations. Two staples of the regimen - clofazimine and single-dose isoniazid - may be difficult to procure in some countries. Moreover, there are no good paediatric formulations of clofazimine and dividing the capsule into smaller doses is impossible, making dosing in children uncertain. The implementation of these guidelines at national level needs to ensure that sufficient quantities of these medicines are available to meet the demand and that no stock-outs occur. The dosing of all medicines in the shorter regimen remain as per the recommendations in the STREAM trial (1).

Direct observation of treatment (DOT) was carried out during the STREAM trial, by clinic staff or family members or other members of the community depending on the local circumstances. Among the deaths in the trial there appears to be an indication that adherence was poor. It is recommended that DOT with patient support be implemented to help patients complete the shorter MDR-TB regimen. In this context the use of digital technologies to support adherence (e.g. video-supported therapy) could have a role as evidence of the effectiveness of this kind of support is strengthened(18). In addition it is recommended that active TB drug-safety monitoring and management (aDSM) be established or utilised in countries implementing the shorter regimen, to detect, manage and report suspected or confirmed adverse events or drug toxicities(19),(20). WHO has published an implementation framework for aDSM which provides additional information on implementation considerations for aDSM. The availability of audiometry services is also important to establish baseline hearing levels and to monitor for hearing loss over time.

If the shorter regimen is used, the GDG recommended that:

- 1. Kanamycin should be replaced by amikacin (on the basis of evidence from the comparative effectiveness of these two injectable agents see PICO question 2)
- 2. Drug susceptibility testing for fluoroquinolones and second-line injectable agents before start of treatment is emphasised, as well as other regimen components where possible (e.g. pyrazinamide, isoniazid mutations)
- 3. Shared decision making between the clinician and patient is important when choosing between a shorter and longer regimen

## **Monitoring and evaluation**

Patients who receive a shorter MDR-TB treatment regimen need to be monitored during treatment using schedules of relevant clinical and laboratory testing which have been successfully applied in the studies under field conditions. If feasible, it is important to follow up patients after the completion of treatment for possible relapse. The STREAM trial (interim) results indicated that relapse occurred in 3.3% in the study arm, which is higher than was inferred from observational studies. However, the final results of the STREAM trial did not demonstrate a statistically significantly higher rate of reversion, relapse or lack of conversion for patients using the shorter regimen.

The WHO framework for aDSM needs to be applied to ensure appropriate action to respond promptly to adverse events and an acceptable level of monitoring for them – alongside the monitoring for treatment outcomes. The use of electrocardiography is still recommended particularly for patients receiving the 800 mg/day dose of moxifloxacin. Audiometry should also be available.

Resistant mutations to fluoroquinolones and second line injectable agents detected using MDRTBs/ should be considered a contraindication for the shorter regimen. Resistance to pyrazinamide (or any other component of the shorter regimen) when determined using reliable DST is also considered an exclusion criterion. However, there is currently no approved rapid test for pyrazinamide susceptibility; given that it may require several weeks to obtain a phenotypic DST result when this is available this test is not imposed as a prerequisite ahead of treatment start. Patients may be started on the shorter regimen until pyrazinamide DST results become available. If a test result eventually shows resistance into treatment with the shorter MDR-TB regimen the clinician may decide to switch to a longer MDR-TB regimen, based on the patient's response to treatment and other considerations.

## Research priorities

While the final results of STREAM Stage 1 have been a welcome development and there is much anticipation for the eventual outcomes of Stage 2 of the trial testing bedaquiline (21). The GDG stressed the importance of continued collection of effectiveness and safety data from cohorts of MDR-TB patients treated with the shorter MDR-TB regimen under observational study conditions. Important variations to the shorter MDR-TB regimen (e.g. removal of prothionamide or ethionamide, or replacement of the injectable agent with bedaquiline to obtain an oral-only regimen) other than those permitted in the trial and observational studies need to be carefully monitored, preferably under operational research conditions(22). Sharing of IPDs from trials and observational studies will remain important to enable pooled analysis of both control and study arms.

GDG members discussed the research priorities for reducing the duration of MDR-TB regimens and highlighted the following priorities:

- · Future research needs to include the effectiveness/safety of the shorter MDR-TB treatment regimen in PLHIV, drug-drug interaction with ARVs, adherence and newer means of supporting adherence.
- The effectiveness/safety of the shorter MDR-TB treatment regimen in subgroups that have been systematically excluded from study protocols (e.g. children, patients with different forms of extrapulmonary disease, pregnant women) and in settings where background resistance to drugs other than fluoroquinolones and second line injectable agents is high (e.g. pyrazinamide or high-level isoniazid resistance).
- · Additional studies of alternative composition of the regimen e.g. replacement of moxifloxacin with levofloxacin; replacement of the injectable with new medicines (already part of STREAM Stage 2); effect of removal of thiamide.
- · Implementation research on the introduction of the shorter MDR-TB regimen, including link with rapid diagnostics for clinical decision-making (e.g. line probe assay).
- · Studies on cost effectiveness (a cost effectiveness analysis is already ongoing as part of the STREAM trial).
- · Studies on aspects related to equity.
- · Studies that include patients with a history of previous MDR-TB treatment.
- · Comparisons of the shorter regimens with high dose versus normal dose fluoroquinolones.
- · Modelling of the population impact of the shorter regimen on the TB epidemic using the inputs from the STREAM trial.
- · Rapid molecular tests for resistance (with information on specific mutations) to correlate with patient outcomes.
- · Possibility to identify patient subgroups in whom the treatment can be reduced to less than 9 months and the impact that this could have on effectiveness and safety.
- · Possibility of a treatment shorter than 9 months with regimens of different composition (as per STREAM Stage 2).

## REFERENCES SUMMARY

- 1. Nunn AJ, Rusen ID, Van Deun A, Torrea G, Phillips PPJ, Chiang C-Y, et al. Evaluation of a standardized treatment regimen of anti-tuberculosis drugs for patients with multi-drug-resistant tuberculosis (STREAM): study protocol for a randomized controlled trial. Trials. 2014;15:353.
- 2. WHO position statement on the use of the shorter MDR-TB regimen (WHO/CDS/TB/2018.2) [Internet]. Geneva, World Health Organization; 2018. Available from: http://www.who.int/tb/publications/2018/
- 3. WHO treatment guidelines for drug-resistant tuberculosis, 2016 update (WHO/HTM/TB/2016.4) [Internet]. Geneva, World Health Organization. 2016. Available from: http://apps.who.int/iris/bitstream/10665/250125/1/9789241549639-eng.pdf
- 4. Preliminary results from STREAM trial provide insight into shorter treatment for multidrug-resistant tuberculosis [Internet]. [cited 2017 Oct 17]. Available from: http://www.ctu.mrc.ac.uk/news/2017/preliminary\_results\_from\_stream\_trial\_provide\_insight\_into\_shorter\_treatment\_for\_multidrug\_resistant\_tuberculosis
- 5. Van Deun A, Maug AKJ, Salim MAH, Das PK, Sarker MR, Daru P, et al. Short, highly effective, and inexpensive standardized treatment of multidrug-resistant tuberculosis. Am J Respir Crit Care Med. 2010 Sep 1;182(5):684–92.
- 6. Kuaban C, Noeske J, Rieder HL, Aït-Khaled N, Abena Foe JL, Trébucq A. High effectiveness of a 12-month regimen for MDR-TB patients in Cameroon. Int J Tuberc Lung Dis. 2015 May;19(5):517–24.
- 7. Piubello A, Harouna SH, Souleymane MB, Boukary I, Morou S, Daouda M, et al. High cure rate with standardised short-course multidrug-resistant tuberculosis treatment in Niger: no relapses. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2014 Oct;18(10):1188–94.
- 8. Trébucq A, Schwoebel V, Kashongwe Z, Bakayoko A, Kuaban C, Noeske J, et al. Treatment outcome with a short multidrug-resistant tuberculosis regimen in nine African countries. Int J Tuberc Lung Dis. 2018 Jan 1:22(1):17–25.
- 9. Global tuberculosis report 2018 (WHO/HTM/TB/2018.20) [Internet]. Geneva, World Health Organization; 2018. Available from: http://apps.who.int/iris/bitstream/handle/10665/274453/9789241565646-eng.pdf
- 10. WHO | Public call for individual patient data on treatment of rifampicin and multidrug-resistant (MDR/RR-TB) tuberculosis [Internet]. WHO. [cited 2018 Feb 27]. Available from: http://www.who.int/tb/features archive/public call treatment RR MDR TB/en/
- 11. Ahmad Khan F, Salim MAH, du Cros P, Casas EC, Khamraev A, Sikhondze W, et al. Effectiveness and safety of standardised shorter regimens for multidrug-resistant tuberculosis: individual patient data and aggregate data meta-analyses. Eur Respir J. 2017 Jul;50(1):1700061.
- 12. Stop TB Partnership | Global Drug Facility (GDF) GDF Products List [Internet]. [cited 2016 Apr 26]. Available from: http://www.stoptb.org/gdf/drugsupply/drugs available.asp
- 13. Trauer JM, Achar J, Parpieva N, Khamraev A, Denholm JT, Falzon D, et al. Modelling the effect of short-course multidrug-resistant tuberculosis treatment in Karakalpakstan, Uzbekistan. BMC Med [Internet]. 2016 Dec [cited 2016 Nov 24];14(1). Available from: http://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-016-0723-2
- 14. Yassin MA, Jaramillo E, Wandwalo E, Falzon D, Scardigli A, Kunii O, et al. Investing in a novel shorter treatment regimen for multidrug-resistant tuberculosis: to be repeated. Eur Respir J. 2017 Mar:49(3):1700081.
- 15. Reuter A. Tisile P. von Delft D. Cox H. Cox V. Ditiu L. et al. The devil we know: is the use of injectable agents for the treatment of MDR-TB justified? Int J Tuberc Lung Dis. 2017 Nov 1:21(11):1114–26.
- 16. Seddon JA, Thee S, Jacobs K, Ebrahim A, Hesseling AC, Schaaf HS. Hearing loss in children treated for multidrug-resistant tuberculosis. J Infect. 2013 Apr;66(4):320–9.
- 17. The use of molecular line probe assays for the detection of resistance to second-line anti-tuberculosis drugs. Policy guidance. (WHO/HTM/TB/2016.07) [Internet]. Geneva, World Health Organization. 2016. Available from: http://www.who.int/tb/publications/lpa-mdr-diagnostics/en/
- 18. Guidelines for the treatment of drug-susceptible tuberculosis and patient care, 2017 update. (WHO/HTM/TB/2017.05) [Internet]. Geneva, World Health Organization. 2017. Available from: http://apps.who.int/iris/bitstream/10665/255052/1/9789241550000-eng.pdf

- 19. Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis. (WHO/HTM/TB/2014.11) [Internet]. Geneva, World Health Organization; 2015. Available from: http://apps.who.int/iris/bitstream/10665/130918/1/9789241548809 eng.pdf
- 20. Active tuberculosis drug-safety monitoring and management (aDSM). Framework for implementation.(WHO/HTM/TB/2015.28) [Internet]. Geneva, World Health Organization; 2015. Available from: http://apps.who.int/iris/bitstream/10665/204465/1/WHO HTM TB 2015.28 eng.pdf
- 21. ISRCTN ISRCTN18148631: STREAM 2 The evaluation of a standard treatment regimen of anti-tuberculosis drugs for patients with multi-drug-resistant tuberculosis [Internet]. Available from: https://doi.org/10.1186/ISRCTN18148631
- 22. Global Drug-resistant TB Initiative (GDI). The evaluation of effectiveness and safety of novel shorter treatment regimens for multidrug-resistant tuberculosis. Operational research protocol template [Internet]. Geneva, Stop TB Partnership; 2018. Available from: http://www.stoptb.org/wg/mdrtb/assets/documents/GDI%20OR%20generic%20protocol%20final.pdf

## **PICO QUESTIONS 2 AND 3**

- 2. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?
- 3. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines are outcomes safely improved with fewer or more than five effective medicines in the intensive phase?

POPULATION:	Patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens conforming to WHO guidelines
INTERVENTION:	An intensive phase with at least five medicines likely to be effective
COMPARISON:	More or less than five medicines likely to be effective
MAIN OUTCOMES:	Treatment failure or relapse versus treatment success; Death versus treatment success; Adverse events (PICO question 2 only)
SETTING:	Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries.
PERSPECTIVE:	The findings of these analyses are expected to influence the continued validity in their present form of recommendations made by WHO in 2016 about the number of medicines likely to be effective, and the choice of individual agents, when composing longer MDR-TB regimens. (for GRADE summary of evidence table see Annex 8).
BACKGROUND:	The previous recommendation for the composition of longer MDR-TB treatment regimens featuring in the WHO Treatment Guidelines for Drug-Resistant Tuberculosis 2016 Update is that patients

should be treated with a regimen with at least five effective medicines during the intensive phase, including pyrazinamide and four second-line TB medicines: one chosen from Group A, one from

Group B, and at least two from Group C (conditional recommendation, very low certainty in the evidence)1. Further, the current recommendation states that if the minimum number of effective TB medicines cannot be composed as above, an agent from Group D2 and other agents from Group D3 may be added to bring the total number of medicines to five. Further, in patients with rifampicin resistant or multidrug-resistant TB it is recommended that the regimen be further strengthened with high-dose isoniazid and/or ethambutol (conditional recommendation, very low certainty in the evidence). Additional interim policy guidance on the use of bedaquiline (published by WHO in 2013) and delamanid (published by WHO in 2014, for use in adults and in 2016, for use in children and adolescents aged 6-17 years) for the treatment of MDR-TB also apply. The number of effective medicines needed to treat MDR-TB is recommended based on the likelihood of the regimen's effectiveness and the need to achieve a good balance of expected benefits to harms. These recommendations have no bearing on the 9-12 month shorter MDR-TB regimen recommended by WHO since 2016 which is standardised in composition and duration.

Given recent developments on the recommendations of new drugs for the treatment of MDR/RR-TB and their increasing use in several settings globally, it was deemed timely to review the composition and the number of effective medicines to be used in longer MDR/RR-TB regimens. The 2018 IPD contains new datasets from several countries (including a comparatively large dataset on the use of bedaquiline for MDR-TB in South Africa from recent years) to inform these questions.

This Evidence to Decision framework combines the information and judgements made on PICO questions 2 (use of individual medicines) and 3 (number of agents likely to be effective):

PICO question 2: to analyse treatment success, failure, relapse and death for the individual medicines in longer regimens, the main individual patient data meta-analysis (IPD-MA) with 13,104 records from 53 studies in 40 countries was used. The individual agents assessed included all agents in the 2016 classification of medicines except for gatifloxacin, high-dose isoniazid and thioacetazone for which too few or no records were available for longer regimens in current use.

PICO guestion 2: to analyse adverse events (AEs) resulting in permanent discontinuation of individual medicines in longer regimens, a subset of 5,450 records from 17 studies in the IPD was used, supplemented with additional information from 10 other studies that only reported AEs for either Bdg (N=130), Lzd (N=508) or carbapenems (N=139).

Separate from these data the GDG also assessed unpublished results from Phase III Trial 213 of delamanid; and safety and pharmacologic exposure data from unpublished paediatric studies of bedaquiline (Phase II TMC207-C211 and Phase I/II IMPAACT P1108) and delamanid (Phase I 242-12-245, Phase I 242-12-232, Phase II 242-07-204, Phase II 242-12-233)

PICO question 3: to analyse treatment success, failure, relapse and death for the optimal number of medicines to include in longer regimens, the data were derived from a subset of 8,957 patients from 47 studies included in the IPD used for PICO question 2 above. Of these, 3,570 patients in 16 studies had information on the start and end dates for individual medicines in which drug susceptibility testing (DST) was reported and 5,387 patients in 31 studies had information on individual medicines used in both the intensive and continuation phases of treatment, as well as DST

<sup>1 &</sup>lt;u>Classification of medicines in the 2016 auidelines</u>: Group A=levofloxacin, moxifloxacin, gatifloxacin; Group B=amikacin, capreomycin, kanamycin, (streptomycin); Group C=ethiona mide (or prothionamide), cycloserine (or terizidone), linezolid, clofazimine; Group D2=bedaquiline, delamanid; Group D3=p-aminosalicylic acid, imipenem-cilastatin, meropenem, amoxicillin-clavulanate, (thioacetazone) [Group D1 was composed of pyrazinamide, ethambutol and high-dose isoniazid]

results. Given that this question focused on the number of agents in both the intensive phase and continuation phases, patients who did not receive an injectable agent or in whom an initial intensive phase was not defined were excluded (N=476). Patients who were designated "cured" or "treatment completed" but received less than 18 months of treatment – the minimum duration for longer regimens - were also excluded (N=346).

For PICO question 3, in cases where DST results were available, a medicine was considered effective if results showed susceptibility and not counted as effective if results showed resistance. Where DST results were missing, two situations existed.

- (1) If the prevalence of resistance to that medicine was <10% in the same population (from same country or study site if within one country, or overall at all sites if local data not available) then the medicine was counted as effective if DST result was missing. This applied to the following agents: cycloserine or terizidone, linezolid, clofazimine, bedaquiline, the carbapenems and delamanid
- (2) If the prevalence of resistance to that medicine was >=10% in the same population (from same country or study site if within one country, or overall at all sites if local data not available) then imputed DST results were used if DST was missing to determine effectiveness. If imputed DST result was susceptible, then the medicine was counted as effective; if imputed DST result was resistant, then the medicine was not counted as effective. This applied to the following agents: pyrazinamide, ethambutol, second line injectable agents, fluoroquinolones, PAS, ethionamide or prothionamide.

The following were not included when counting the number of medicines likely to be effective (regardless of any DST result which may have been available): isoniazid (including high-dose isoniazid), rifampicin, rifabutin, thioacetazone, amoxicillin-clavulanate or macrolide antibiotics.

#### **ASSESSMENT**

Problem Is the problem a priority?							
JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS					
o No o Probably no o Probably yes ● Yes o Varies o Don't know	About 558,000 new cases of rifampicin- and multidrug-resistant TB (MDR/RR-TB) cases are estimated to emerge worldwide each year and 230,000 cases die. Globally only about one fourth of newly emergent MDR-TB patients have been reported to start a second-line TB treatment annually in recent years.  Outcomes of MDR-TB treatment on a global level are poor with much loss to follow up and death, and only 55% of cases reported to have a successful outcome at the end of treatment. The need for better treatment for these patients is thus a priority in an effort to save lives and reduce transmission and chronicity.	MDR-TB is a global challenge and access to treatment often problematic, with regimens typically being long, toxic and expensive.  There is clearly an interest in reducing the duration of treatment, simplifying the administration of the treatment regimen, and providing patients with a safer combination of medicines that can cure the large majority of cases.					

# **Desirable Effects**

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE					ADDITIONAL CONSIDERATIONS		
o Trivial o Small	Table 1. Summary of	f findings on the	number of effecti	The likelihood of treatment success in MDR-TB patients on longer regimens depends upon patient/strain level factors (including severity of disease,				
o Moderate o Large • Varies o Don't know	Outcomes	INTERVENTION COMPARA		Difference Relative effect (95% CI)		resistance patterns and co-morbidities) as well as access to health care ( regimens with sufficient effective agents, medications of good quality, attention to adverse events and patient support). Longer MDR-TB regim with sufficient effective agents are known to increase the likelihood of c		
	Treatment Failure/Relapse vs. Treatment Success: 4 Possibly Effective Drugs (INTERVENTION) vs. 5 Possibly Effective Drugs in the Intensive Phase (COMPARATOR)	13 per 100	<b>13 per 100</b> (9 to 17)	O fewer per 100 (4 fewer to 4 more)	<b>aOR 1.0</b> (0.7 to 1.3)	and lower the risk of death in adults and children (1),(2),(3),(4).  The GDG assessed the findings of the analysis for the number of medicines to be used in the intensive and continuation phase of the regimens by comparing the risk for treatment failure and relapse with success and death with treatment success. The analysis showed that contemporary longer MDR-TB treatment regimens containing 4 agents likely to be effective in the intensive phase had a similar likelihood of favourable outcomes to regimens starting with 5 or 6 effective agents (Table 1). The analysis also showed that patients with 3 agents in the continuation phase - the situation expected when starting		
	Death vs. Treatment Success: 4 Possibly Effective Drugs (INTERVENTION) vs. 5 Possibly Effective Drugs in the Intensive Phase (COMPARATOR)	19 per 100	<b>21 per 100</b> (18 to 29)	2 more per 100 (2 fewer to 10 more)	<b>aOR 1.1</b> (0.9 to 1.5)	with 4 agents and stopping the injectable agent at the end of the intensive phase - fared no worse than those with 4 agents. Given that the likelihood of AEs, drug-drug interactions and pill burden increases with the number of agents included in a regimen it would be desirable to give patients the minimum number of medicines necessary to obtain comparable levels of relapse-free cure (see also Table 2 under Undesirable effects).  The GDG also assessed the individual contribution to patient outcomes of medicines used in longer MDR-TB regimens using primarily the estimates of		
	Treatment Failure/Relapse vs. Treatment Success: 6 Possibly Effective Drugs (INTERVENTION) vs. 5 Possibly Effective Drugs in the Intensive Phase (COMPARATOR)	10 per 100	<b>10 per 100</b> (6 to 17)	O fewer per 100 (4 fewer to 7 more)	<b>aOR 1.0</b> (0.6 to 1.7)	effect from the IPD MA and Trial 213 (delamanid) to address PICO question 2 (see GRADE SOF tables for each medicine in Annex 8 of the guidelines; summary of judgements in Table 3 further down). Following a thorough assessment of benefits to harms the medicines were classified in three groups (see Table 4 under Balance of Effects). Three agents were considered highly effective (fluoroquinolones, bedaquiline and linezolid; Group A), and cycloserine/terizidone and clofazimine to be preferred second choice regimen components (Group B). The rest of the medicines were assigned to Group C, for use if the regimen cannot be composed with Group A and B agents alone. Judgements on the individual medicines in the regimen are summarised in Table 3 (under Balance of Effects). As to the other medicines, Km and Cm were		
	Death vs. Treatment Success: 6 Possibly Effective Drugs (INTERVENTION) vs. 5 Possibly Effective Drugs in the Intensive Phase (COMPARATOR)	ys. e (7 to 14)	1 fewer per 100 (4 fewer to 3 more)	<b>aOR 0.9</b> (0.6 to 1.3)	associated with poorer outcomes; Gfx, T and high-dose H were rarely used in contemporary longer regimens; and perchlozone, interferon gamma and sutezolid could not be considered further owing to the absence of <i>final</i> treatment outcome data from appropriate patient studies (i.e. no recommendation would be possible).  The GDG discussed how to deal with oral-only regimens in which an injectable agent was not given in the first months, which has up to now defined the intensive phase (including for the 2018 IPD MA). Pinning a recommendation to			

Treatment	17 per 100	22 per 100	5 more per 100	aOR 1.3
Failure/Relapse vs. Treatment Success: 2 Possibly Effective Drugs (INTERVENTION) vs.		(13 to 35)	(3 fewer to 18 more)	(0.8 to 2.1)
3 Possibly Effective Drugs in the Continuation Phase (COMPARATOR)				
Death vs. Treatment Success: 2 Possibly Effective Drugs (INTERVENTION) vs. 3 Possibly Effective Drugs in the Continuation Phase (COMPARATOR)	20 per 100	<b>26 per 100</b> (16 to 39)	6 more per 100 (4 fewer to 20 more)	<b>aOR 1.3</b> (0.8 to 2.0)
Treatment Failure/Relapse vs. Treatment Success: 4 Possibly Effective Drugs (INTERVENTION) vs. 3 Possibly Effective Drugs in the Continuation Phase (COMPARATOR)	10 per 100	<b>12 per 100</b> (9 to 15)	2 more per 100 (1 fewer to 5 more)	<b>aOR 1.2</b> (0.9 to 1.5)
Death vs. Treatment Success: 4 Possibly Effective Drugs (INTERVENTION) vs. 3 Possibly Effective Drugs in the Continuation Phase (COMPARATOR)	13 per 100	<b>13 per 100</b> (11 to 18)	<b>0 fewer per 100</b> (3 fewer to 4 more)	<b>aOR 1.0</b> (0.8 to 1.3)

application to regimens containing an injectable agent. When making the recommendation on the number of effective agents the GDG was also mindful of the need to provide for situations in which a medicine is only used in the first months because of indication (i.e. Bdq and Dlm, that would normally be stopped 6 months after start) or because of tolerability (especially Lzd), meaning that for most of its duration the regimen would contain one or two key agents less than at the start.

In conclusion, the GDG considered that in patients treated with Group A and Group B agents alone, starting the regimen with 4 agents likely to be effective and continuing with 3 once Bdq is stopped at 6 months would be a suitable regimen for most patients. If the regimen cannot be composed of Group A and B agents alone additional agents from Group C may be needed.

Given that the regimen needs to have at least 3 effective agents after Bdq is stopped at 6 months, if another agent needs to be stopped because of toxicity then that medicine is replaced by another one. The replacement medicine would be chosen either from Group B (unless both Cfz and Cs/Trd are already included) or from Group C. The choice from Group C is determined by the order in which the medicines are ranked and the individual circumstances of the patient and setting.

Starting with 5 agents instead of 4 may be favoured in certain situations to avoid the need to replace a medicine into treatment, namely:

- two of the four agents are likely to be stopped before the end of treatment, for instance Bdq stopped at month 6 and Lzd stopped early because of toxicity;
- (ii) reliable DST is not available for one or more of the agents on the regimen but background resistance to the agent is known to be high;
- (iii) the agents included in the regimen are unlikely to cure the patient (e.g. only a total of 2 of the agents from Group A and Group B are included in the regimen)

The GDG members considered it very important that the new recommendations on regimen design are accompanied by continued advocacy for greater access to DST, both for medicines to which reliable methods exist as well as others in development for newer medicines.

#### **Undesirable Effects**

How substantial are the undesirable anticipated effects?

#### JUDGEMENT

#### RESEARCH EVIDENCE

# o Large

- o Moderate o Small
- VariesO Don't know

o Trivial

# Table 2. Serious adverse events (SAEs) in patients on longer MDR-TB regimens (PICO 2)

뷕

F1				
Medicine	Abso	lute risk of AE		
	Median %	95% credible interval		
Bedaquiline	2.4%	[0.7, 7.6]		
Moxifloxacin	2.9%	[1.4, 5.6]		
Amoxycillin-Clavulanic acid	3.0%	[1.5, 5.8]		
Clofazimine	3.6%	[1.3, 8.6]		
Ethambutol	4.0%	[2.4, 6.8]		
Levofloxacin	4.1%	[1.9, 8.8]		
Streptomycin	4.5%	[2.3, 8.8]		
Cycloserine / terizidone	7.8%	[5.8, 10.9]		
Capreomycin	8.4%	[5.7, 12.2]		
Pyrazinamide	8.8%	[5.6, 13.2]		
Ethionamide / prothionamide	9.5%	[6.5, 14.5]		
Amikacin	10.3%	[6.6, 17.0]		
Kanamycin	10.8%	[7.2, 16.1]		
p-aminosalicylic acid	14.3%	[10.1, 20.7]		
Thiacetazone	14.6%	[4.9, 37.6]		
Linezolid	17.2%	[10.1, 27.0]		

**NOTE:** From an "arm-based network" meta-analysis of a patient subset from the 2016 IPD for which adverse events resulting in permanent discontinuation of a TB medicine (27 studies) or classified as Grade 3-5 (3 studies) were reported. There were insufficient records on delamanid, imipenem-cilastatin and meropenem to estimate risks. Agents that are not in Groups A, B or C are italicized.

The evidence included assessments of the efficacy of individual medicines as components of a longer regimen, assessing their effects on the outcomes of treatment failure/ relapse and death (compared to treatment success) and adverse events. While the estimates of effect varied for each individual medicine assessed and by outcome, the most effective medicines have been categorised and ranked in a revised grouping of TB medicines recommended for use in the longer MDR-TB regimens, regrouped into three categories (i.e. Groups A-C), based on the evidence about the balance of effectiveness to safety. The GRADE tables in Annex 8 of the guidelines summarise the effectiveness of each individual medicine relative to the outcomes of interest and the number of effective medicines required in the intensive phase, as part of the longer regimen.

#### ADDITIONAL CONSIDERATIONS

Just as for benefit from treatment, the likelihood of harms from MDR-TB treatment is expected to vary depending on patient factors (e.g. disease severity, comorbidities) and the health intervention (choice of medicines, pill burden and drug-drug interactions, adequacy of safety monitoring and support, options to switch drugs in case of adverse reactions) (Table 2). The more medicines are used the higher is the likelihood of adverse events and drug-drug interactions. In addition, the GDG noted that among the possibly effective medicines included in the IPD, injectable agents that are now deemed not highly effective (based on the new analyses conducted for PICO question 2) were included.

Based on the new regrouping of medicines (Table 4), some medicines associated with serious adverse reactions - particularly kanamycin, capreomycin and thioacetazone - are now no longer included among the components of the longer MDR-TB regimen. Lzd has now become a Group A drug that is strongly recommended and may be used more frequently than it was in the past although it has the highest likelihood of serious adverse events of all the second line medicines studied. On this basis the GDG considered that the benefit of inclusion of Lzd would outweigh the risks. Use of Lzd for at least 6 months was shown to increase effectiveness, although toxicity may limit use. The analysis suggested that using Lzd for the whole duration of treatment would optimise its effect (about 70% of patients on Lzd with data received it for more than 6 months and 30% for 18 months or the whole duration). No patient predictors for early cessation of Lzd could be inferred from the IPD sub-analysis. Nonetheless it is expected that in a number of situations this medicine may need to be replaced later on in the treatment because of safety concerns. Amikacin and the carbapenems, that showed good effectiveness in the IPD MA, were ranked lower in the priority listing because of the associated inconvenience of daily injection and - for amikacin - the risk of ototoxicity and nephrotoxicity in particular. The risk of SAE for patients on delamanid could not be assessed owing to the sparsity of records for patients on this medicine in the IPD; however Trial 213 results attested to the relatively good safety record of delamanid in adults with pulmonary forms of disease(5).

A number of longer regimen components are associated with QT-interval prolongation (moxifloxacin, bedaquiline, clofazimine, delamanid) and their simultaneous use may increase this risk. Levofloxacin may be preferred over moxifloxacin to limit the risk of additive cardiotoxicity.

The additional data on safety and pharmacology of Bdq in patients <18 years and for Dlm in children <6 years of age were assessed by the GDG (full report at Annex 10 of guidelines). Operating under the assumption that exposure-response (efficacy) profiles can be extrapolated from adults to children aged 3 years and more, the GDG concluded that:

For Bdq, the safety risk in children down to 6 years of age as shown from the

trial population does not appear to exceed that of adults. However, the variability present in the limited sample size precluded a comment on exposure-response (safety). Nonetheless, the doses evaluated do not appear to produce exposures that would put children at increased risk for therapeutic failure.

For *Dlm*, the focus was children aged 3-5 years, given that a WHO recommendation for use in 6-17 year olds had already been formulated in 2016(6), while exposures at the three doses explored in trials of children under 3 years fell below adult values. The exposure profiles in 3-5 year olds were comparable to adults and no higher than in children 6 years of age and older whose data had already been assessed for the 2016 recommendation. Moreover, no safety signals distinct from those reported in adults were observed in 3-5 year olds. The GDG nonetheless had concerns about the feasibility of administering the correct dose to children aged 3-5 years given that the special formulation used in the trial (25mg) was not available and only the adult tablet exists (50mg), which is not bioequivalent and presents challenges to manipulate its contents without compromising its effectiveness.

In its judgements the GDG highlighted the importance of proper attention to the monitoring of adverse events (aDSM), and that special investigations like audiometry and electrocardiography be put in place as needed. Considerations for the use of the 9-12 month shorter regimen in eligible patients would also depend upon an assessment of likelihood of benefits and harms should such patients be given an individualised longer regimen instead. The GDG underlined the importance of DST to avoid including ineffective agents that can add to the toxicity and that could be replaced by more effective agents.

Table 3. Summary of judgements and strength of recommendations for individual components of longer MDR-TB regimens (PICO 2)

	Group A		Group A Group B Group C						Other medicines									
Judgement	levofloxacin or moxifloxacin	bedaquiline	linezolid	clofazimine	cycloserine or terizidone	ethambutol	delamanid	pyrazinamide	carbapenems (imipenem- cilastatin / meropenem)	amikacin	streptomycin	prothionamide or ethionamide	p - aminosalicylic acid (PAS)	kanamycin and capreomycin	amoxicillin- clavulanic acid	thiacetazone	isoniazid	gatifloxacin
Balance of effects	Favors the intervention	Favors the intervention	Favors the intervention						Probably favors the intervention	Probably favors the intervention		Varies	Probably favors the comparison		Favors the comparison	Not assessed - no data	Not assessed - insufficient data	Not assessed - insufficient data
Certainty of evidence	Moderate	Moderate	Moderate	Very low	Very low	Very low	Low	Very low	Very low	Very low	Very low	Very low	Very low	Very low	Low	Not assessed - no data	Not assessed - insufficient data	Not assessed - insufficient data
Resources required	Varies	Varies	Varies	Negligible costs and savings	Moderate costs	Moderate costs	Varies	Negligible costs and savings	Large costs	Large costs	Large costs	Negligible costs and savings	Moderate costs	Large costs	Negligible costs and savings		Not assessed - insufficient data	Not assessed - insufficient data
Cost effectiveness	No included studies	No included studies	No included studies	No included studies	No included studies	No included studies	No included studies	No included studies	No included studies	No included studies	No included studies	No included studies	No included studies	No included studies	No included studies	Not assessed - no data	No included studies	No included studies
Equity	Probably increased	Probably increased	Probably increased	Probably no impact	Probably no impact	Probably no impact	Varies	Probably no impact	Probably reduced	Probably no impact	Probably no impact	Probably no impact	Probably no impact	Reduced	Probably no impact	Not assessed - no data	Not assessed - insufficient data	Not assessed - insufficient data
Acceptability	Yes	Yes	Probably yes	Probably yes	Probably yes	Probably yes	Probably yes	Probably yes	Probably no	Probably no	Probably no	Probably no	Probably no	No	No	Not assessed - no data	Not assessed - insufficient data	Not assessed - insufficient data
Feasibility	Yes	Probably yes	Probably yes	Probably yes	Probably yes	Probably yes	Varies	Yes	Varies	Probably yes	Varies	Probably yes	Probably yes	Probably yes	Yes	Not assessed - no data	Not assessed - insufficient data	Not assessed - insufficient data
Strength of recommendation	Strong	Strong	Strong	Conditional	Conditional	Conditional	Conditional	Conditional	Conditional	Conditional	Conditional	Conditional	Conditional	Conditional (against)	Strong (against use without Imp-Cln or Mpm)	-	-	-

# Certainty of evidence What is the overall certainty of the evidence of effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low  ● Low o Moderate o High o No included studies	The primary source of evidence for PICO question 2 is a pooled individual patient dataset of 53 studies from 40 countries composed of 13,104 patients with final outcomes of treatment (main 2018 IPD LR MA). Of the 53 studies, 50 were cohort studies and three randomised controlled trials. Twenty percent of the patients in the IPD-MA were from high income countries, and 58% and 22% were from middle and low-income countries respectively. Trial 213 data were available for delamanid and other data on pharmacology and safety of Bdq and Dlm in children were also available. Data on adverse events were derived from a separate IPD MA for 30 observational studies. For PICO question 3, a subset of the 2018 IPD LR MA was used, comprising 47 studies and 8957 patients. Of these, information on the number of possibly effective drugs used in the intensive phase was available for 8135 patients and for the continuation phase for 6883. Overall, the certainty of the evidence was classified as "low", which has improved from the "very low" certainty attached to the data sources used for the 2011 and 2016 guidelines updates.	The certainty of the evidence is low based on the fact that the majority of studies included in the IPD-MA are observational. The analyses of the 2018 IPD data were matched (exactly) for resistance to a fluoroquinolone, resistance to a second line injectable agent and income level (according to the World Bank Atlas method). In addition, propensity score matching was used to adjust for a number of other important covariates. However, while these efforts compensated for some potential confounders, residual bias is likely to be present.

### Values

Is there important uncertainty about or variability in how much people value the main outcomes?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Important uncertainty or variability o Possibly important uncertainty or variability o Probably no important uncertainty or variability • No important uncertainty or variability	No research evidence was identified.	Although no research evidence was identified, the GDG considered that there was no important uncertainty or variability in how much people would value the outcomes that were assessed (treatment success, failure, relapse, death and adverse events).

#### **Balance of effects**

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

#### JUDGEMENT

#### RESEARCH EVIDENCE

# The likelihood of desirable effects (success) and undesirable effects (treatment failure, relapse, death or adverse events) was estimated for individual medicines. The same outcomes (except adverse events) were also calculated for combinations of 4, 5 and 6 medicines (see GRADE Summary of Findings tables in Annex 8 of the guidelines)

#### Table 4. Grouping of medicines recommended for use in longer MDR-TB regimens [1]

comparison
O Does not favor either the
intervention or the
comparison
o Probably favors the
intervention
o Favors the intervention
<ul><li>Varies</li></ul>
o Don't know

o Favors the comparison

o Probably favors the

GROUPS & STEPS	MEDICINE	
Group A: Include all three medicines	Levofloxacin <u>OR</u> Moxifloxacin	Lfx Mfx
Therade un three medicines	Bedaquiline <sup>2,3</sup>	Bdq
	Linezolid <sup>4</sup>	Lzd
Group B:	Clofazimine	Cfz
Add one or both medicines	Cycloserine <u>OR</u> Terizidone	Cs <u>Trd</u>
Group C:	Ethambutol	Е
Add to complete the regimen and when	Delamanid <sup>3,5</sup>	Dlm
medicines from Groups A and B cannot be	Pyrazinamide <sup>6</sup>	Z
used	Imipenem- <u>cilastatin</u> OR Meropenem <sup>7</sup>	Ipm-Cln Mpm
	Amikacin (OR Streptomycin) <sup>8</sup>	Am (S)
	Ethionamide <u>OR</u> Prothionamide <sup>9</sup>	Eto Pto
	p-aminosalicylic acid9	PAS

- 1. This table is intended to guide the design of individualized, longer MDR-TB regimens (the composition of the recommended shorter MDR-TB regimen is largely standardized; see Section A). Medicines in Group C are ranked by decreasing order of usual preference for use subject to other considerations (see text for more details). The 2018 IPD-MA for longer regimens included no patients on thioacetazone (T) and too few patients on gatifloxacin (Gfx) and high-dose isoniazid (Hh) for a meaningful analysis. No recommendation on perchlozone, interferon gamma or sutezolid was possible based on patient treatment outcome data currently available.
- 2. Evidence on the safety and effectiveness of Bdq beyond 6 months and below the age of 6 years was insufficient for review. Use of Bdq beyond these limits should follow best practices in 'off-label' use.
- 3. Evidence on the concurrent use of Bdg and Dlm was insufficient for review.
- 4. Use of Lzd for at least 6 months was shown to increase effectiveness, although toxicity may limit use. The analysis suggested that using Lzd for the whole duration of treatment would optimise its effect (about 70% of patients on Lzd with data received it for more than 6 months and 30% for 18 months or the whole duration). No patient predictors for early cessation of Lzd could be inferred from the IPD sub-analysis.
- 5. Evidence on the safety and effectiveness of Dlm beyond 6 months and below the age of 3 years was insufficient for review. Use

#### ADDITIONAL CONSIDERATIONS

The composition of a longer MDR-TB regimen varies in both which medicines are used and how many are combined. These two considerations, along with duration of use (see PICO questions 4-6), have an important bearing on both the desirable and undesirable effects. When making judgements on regimen composition the GDG expects that the balance of effects will probably favour the intervention in most instance but acknowledges that this may vary when options are few and the clinician is compelled to treat with regimens for which certainty of effect is very low or backed by no data.

Overall, the risk-benefit considerations for the use of Bdq in patients aged 6-17 years and of Dlm in children aged 3-5 years would be similar to those considered for adults.

of Dlm beyond these limits should follow best practices in 'off-label' use.

- 6. Z is only counted as an effective agent when DST results confirm susceptibility.
- 7. Every dose of Imp-Cln and Mpm is administered with clavulanic acid, which is only available in formulations combined with amoxicillin (Amx-Clv). Amx-Clv is not counted as an additional effective TB agent and should not be used without Imp-Cln or Mpm.
- 8. Am and S are only to be considered if DST results confirm susceptibility and high-quality audiology monitoring for hearing loss can be ensured. S is to be considered only if Am cannot be used and if DST results confirm susceptibility (S resistance is not detectable with 2nd line molecular line probe assays and phenotypic DST is required). Kanamycin (Km) and capreomycin (Cm) are no longer recommended for use in MDR-TB regimens.
- 9. These agents only showed effectiveness in regimens without Bdq, Lzd, Cfz or Dlm, and are thus only proposed when other options to compose a regimen are not possible.

# **Resources required**

How large are the resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Large costs o Moderate costs o Negligible costs and savings o Moderate savings o Large savings ● Varies o Don't know	No research evidence was identified.	Although no research evidence was identified, the GDG had access to current unit costs of the medicines as according to the Products Catalogue of the Global Drug Facility of the Stop TB Partnership, a major provider of TB medicines in low-income countries (http://www.stoptb.org/gdf/drugsupply/drugs_available.asp; prices exclusive of expenses needed for packaging, delivery and custom clearance at port of arrival). On the basis of these prices the cost of a new longer MDR-TB regimens composed of Group A and B agents alone is expected to increase the cost of a current regimen by about 3 times, but this price could vary up or down if certain medicines have to be replaced with Group C agents.  The GDG also remarked that important drug price reductions have occurred in recent years, especially for Mfx and Lzd, making them more affordable. Moreover, a Bdq donation programme, and ongoing negotiations to bring down the price of this agent for low resource settings after the end of the donation programme in March 2019, will make this medicine more accessible.  Other policies on patient hospitalisation and monitoring are expected to contribute significantly to the resources needed to implement the new regimens. For instance, decentralising care and limiting hospitalization to patients needing inpatient treatment could cut down on service and patient costs.

# **Certainty of evidence of required resources**What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low o Low o Moderate o High • No included studies	No research evidence was identified.	The GDG considered that there were too many variables that could influence resource use. Without a dedicated study it is not easy to quantify the swings that may occur when certain policies are used over others (e.g. hospitalization to treat complications of the disease or adverse events may be expected to raise costs in all settings but by a different amount depending on country)

# **Cost effectiveness**

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Favors the comparison	No research evidence was identified.	CEAs for bedaquiline and delamanid have been conducted in the past
<ul> <li>Probably favors the</li> </ul>		but were not considered at length by the GDG because the base
comparison		assumptions may be outdated or of limited applicability (modelled
O Does not favor either the		pricing, magnitude of expected added benefit, position in regimens as
intervention or the		replacement/add-on, most studies in high income settings etc). The
comparison		GDG did not put much value on CEAs for policy making of MDR-TB
<ul> <li>Probably favors the</li> </ul>		treatment, given that options to design an effective and safe
intervention		treatment are often limited and decision-makers are rarely faced with
<ul> <li>Favors the intervention</li> </ul>		the luxury of having to choose between multiple treatments of equal
o Varies		effectiveness in which the decision hinges mainly upon the cost of the
<ul> <li>No included studies</li> </ul>		regimen and its delivery.

**Equity**What would be the impact on health equity?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Reduced o Probably reduced o Probably no impact o Probably increased o Increased ● Varies o Don't know	No research evidence was identified.	Although no research evidence was identified, the GDG determined that the impact on heath equity may vary depending on the context. For example, in settings with finite resources, access to the newer medicines may be limited, at least initially. Purchasing more expensive treatments without increasing the budget may mean that less patients can receive the most effective/safe options in care. However, if these medicines become cheaper and available to more patients over the medium to long term then this may increase health equity.

# Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
O No O Probably no ● Probably yes O Yes O Varies O Don't know	No research evidence was identified.	Although no research evidence was identified, the GDG acknowledged that many existing regimens for MDR/RR-TB use combinations of all medicines recommended in the newer regimen to varying extents. However, the use of the newer medicines and their combination as is now being recommended is less common in many high burden countries so far and may therefore be less acceptable to clinicians until they become more familiar with them. All the medicines in Groups A to C are otherwise marketed and available (including via the GDF catalogue). The GDG also concluded that an all oral regimen for MDR/RR-TB would very quickly become more acceptable to most patients and health care workers, should all medicines be available and if they are fully aware of the added benefits. Given that the new recommendation favours regimens that start with less medicines the likelihood of pill burden and adverse drug reactions is reduced further. Political commitment and donor support to ensure smooth transition (e.g. extra expense to introduce newer regimens and the need at times to dispose of stocks of older medicines) would be important. The main advantage of the shorter MDR-TB regimen for patients who are eligible is that it is half as long as the longer regimen but comes with the inconvenience of daily intramuscular injections with Am for at least 4 months and a slightly lower likelihood of success (see also PICO question 1).

# **Feasibility**

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o No o Probably no ● Probably yes o Yes o Varies o Don't know	No research evidence was identified.	Although no research evidence was identified, the GDG acknowledged that the longer MDR-TB regimens used at a large scale globally in the last decade used many of the medicines that are now being recommended. However, the increased use of the new medicine bedaquiline and others that have been less widely used such as linezolid and clofazimine may be challenged by in-country preparations for use (e.g. registration or waivers, procurement mechanisms, national policies and clinician resistance). Countries may also need to increase their budgets to account for the newly recommended composition of the longer MDR-TB regimens. As has been the experience in the past when changes were introduced the feasibility of providing the newly recommended longer MDR-TB regimen should improve over time once these initial implementation and logistical challenges are overcome. For certain medicines the options to administer treatment as recommended may also be a challenge, such as to achieve the correct dose of Dlm in children under 6 years of age. The GDG was encouraged by the global expansion of molecular resistance testing and DST for both first-line and second-line TB medicines and expects that this global trend will continue and could facilitate the introduction of tests for newer medicines in future taking advantage of existing platforms and infrastructure.

# **SUMMARY OF JUDGEMENTS**

	JUDGEMENT							
PROBLEM	No	Probably no	Probably yes	Yes		Varies	Don't know	
DESIRABLE EFFECTS	Trivial	Small	Moderate	Large		Varies	Don't know	
UNDESIRABLE EFFECTS	Large	Moderate	Small	Trivial		Varies	Don't know	
CERTAINTY OF EVIDENCE	Very low	Low	Moderate	High			No included studies	
VALUES	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability				
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	Don't know	
RESOURCES REQUIRED	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know	
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	Low	Moderate	High			No included studies	
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	No included studies	
EQUITY	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know	
ACCEPTABILITY	No	Probably no	Probably yes	Yes		Varies	Don't know	
FEASIBILITY	No	Probably no	Probably yes	Yes		Varies	Don't know	

#### TYPE OF RECOMMENDATION

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	Strong recommendation for the intervention
0	0	0	•	0

#### **CONCLUSIONS**

#### Recommendation

In MDR/RR-TB patients on longer regimens, all three Group A agents and at least one Group B agent should be included to ensure that treatment starts with at least four TB agents likely to be effective and that at least three agents are included for the rest of treatment after bedaquiline is stopped[1]. If only one or two Group A agents are used, both Group B agents are to be included. If the regimen cannot be composed with agents from Groups A and B alone, Group C agents are added to complete it (conditional recommendation, very low certainty in the estimates of effect).

#### **Group A medicines:**

- Levofloxacin or moxifloxacin and linezolid should be included in the treatment of MDR/RR-TB patients on longer regimens (strong recommendation, moderate certainty in the estimates of effect).
- Bedaquiline should be included in longer MDR-TB regimens for patients aged 18 years or more (strong recommendation, moderate certainty in the estimates of effect). Bedaquiline may also be included in longer MDR-TB regimens for patients aged 6-17 years (conditional recommendation, very low certainty in the estimates of effect).

#### **Group B medicines:**

- Clofazimine and cycloserine or terizidone may be included in the treatment of MDR/RR-TB patients on longer regimens (conditional recommendation, very low certainty in the estimates of effect).

#### **Group C medicines:**

- Ethambutol may be included in the treatment of MDR/RR-TB patients on longer regimens (conditional recommendation, very low certainty in the estimates of effect).
- Delamanid may be included in the treatment of MDR/RR-TB patients aged 3 years or more on longer regimens (conditional recommendation, moderate certainty in the estimates of effect).
- Pyrazinamide may be included in the treatment of MDR/RR-TB patients on longer regimens (conditional recommendation, very low certainty in the estimates of effect).
- Imipenem-cilastatin or meropenem may be included in the treatment of MDR/RR-TB patients on longer regimens (conditional recommendation, very low certainty in the estimates of effect). [2]
- Amikacin may be included in the treatment of MDR/RR-TB patients aged 18 years or more on longer regimens when susceptibility has been demonstrated and adequate measures to monitor for adverse reactions can be ensured. If amikacin is not available, streptomycin may replace amikacin under the same conditions (conditional recommendation, very low certainty in the estimates of effect).
- Ethionamide or prothionamide may only be included in the treatment of MDR/RR-TB patients on longer regimens if bedaquiline, linezolid, clofazimine or delamanid are not used or if better options to compose a regimen are not possible (conditional recommendation against use, very low certainty in the estimates of effect).
- p-aminosalicylic acid may only be included in the treatment of MDR/RR-TB patients on longer regimens if bedaquiline, linezolid, clofazimine or delamanid are not used or if better options to compose a regimen are not possible (conditional recommendation against use, very low certainty in the estimates of effect).

#### Other medicines: [3]

- Kanamycin and capreomycin are not to be included in the treatment of MDR/RR-TB patients on longer regimens (conditional recommendation, very low certainty in the estimates of effect).
- Clavulanic acid should not be included in the treatment of MDR/RR-TB patients on longer regimens (strong recommendation against use, low certainty in the estimates of effect). [2]

#### Notes

- [1] Group A = levofloxacin/moxifloxacin, bedaquiline, linezolid; Group B = clofazimine, cycloserine/terizidone; Group C = ethambutol, delamanid, pyrazinamide, imipenem-cilastatin, meropenem, amikacin (streptomycin), ethionamide/prothionamide, p-aminosalicylic acid (see also Table 1 above).
- [2] Imipenem-cilastatin (Imp-Cln) and meropenem (Mpm) are administered with clavulanic acid, which is only available in formulations combined with amoxicillin (Amx-Clv). Amx-Clv is not counted as an additional effective TB agent and should not be used without Imp-Cln or Mpm.
- [3] The 2018 IPD-MA for longer regimens included no or few records of patients treated with gatifloxacin (Gfx), thioacetazone (T) and high-dose isoniazid (Hh) and therefore no recommendations were made on the use of these agents. Quality-assured preparations of Gfx and T are not currently available. The performance of high-dose isoniazid in children underpinned the 2016 recommendation for its use in adults and children; http://apps.who.int/iris/bitstream/10665/250125/1/9789241549639-eng.pdf). Owing to the absence of final treatment outcome data from appropriate patient studies the GDG considered that no recommendation on perchlozone, interferon gamma or sutezolid was possible.

#### **Justification**

Given the recent release of new results from trials and the continued accrual of observational study data for MDR/RR-TB patients on treatment, it was deemed timely to review the composition and the number of effective medicines to be used in longer MDR/RR-TB regimens. The 2018 individual patient data longer regimen meta-analysis (2018 IPD LR MA) contains new datasets from several countries (including a comparatively large dataset on the use of bedaquiline for MDR-TB in South Africa from recent years) to inform these questions.

This Evidence to Decision framework combines the information and judgements made on PICO questions 2 (use of individual medicines) and 3 (number of agents likely to be effective):

PICO question 2: to analyse treatment success, failure, relapse and death for the individual medicines in longer regimens, the main individual patient data meta-analysis (IPD-MA) with 13,104 records from 53 studies in 40 countries was used.

PICO question 2: to analyse adverse events (AEs) resulting in permanent discontinuation of individual medicines in longer regimens, a subset of 5,450 records from 17 studies in the IPD was used, supplemented with additional information from 10 other studies that only reported AEs for either Bdq (N=130), Lzd (N=508) or carbapenems (N=139).

Separate from this the GDG also assessed unpublished results from Phase III Trial 213 of delamanid; and safety and pharmacologic exposure data from unpublished paediatric studies of bedaquiline (Phase II TMC207-C211 and Phase I/II IMPAACT P1108) and delamanid (Phase I 242-12-245, Phase I 242-12-232, Phase II 242-07-204, Phase II 242-12-233)

PICO question 3: to analyse treatment success, failure, relapse and death for the optimal number of medicines to include in longer regimens, the data were derived from a subset of 8,957 patients from 47 studies included in the IPD used for PICO question 2 above. Of these, 3,570 patients in 16 studies had information on the start and end dates for individual medicines in which drug susceptibility testing (DST) was reported and 5,387 patients in 31 studies had information on individual medicines used in both the intensive and continuation phases of treatment, as well as DST results. Given that this question focused on the number of agents in both the intensive phase and continuation phases, patients who did not receive an injectable agent or in whom an initial intensive phase was not defined were excluded (N=476). Patients who were designated "cured" or "treatment completed" but received less than 18 months of treatment – the minimum duration for longer regimens - were also excluded (N=346).

For PICO question 3, a medicine was considered possibly effective if:

- 1. There were documented DST results for the following medicines with an overall prevalence of resistance of ≥10% in the 2018 IPD, i.e. for pyrazinamide, ethambutol, second line injectable agents, fluoroquinolones, PAS, ethionamide or prothionamide. If DST was unavailable for these medicines, the imputed resistance was used; or
- 2. There was documented susceptibility on DST; or
- 3. No DST results and susceptibility presumed for medicines for which the overall prevalence of resistance was <10%, i.e. for cycloserine or terizidone, linezolid, clofazimine, bedaquiline, the carbapenems and delamanid. The following were not included when counting the number of medicines likely to be effective (regardless of any DST result which may have been available): isoniazid (including high-dose isoniazid), rifampicin, rifabutin, thioacetazone, amoxicillin-clavulanate or macrolide antibiotics.

## **Subgroup considerations**

MDR/RR-TB alone or with additional resistance: A longer regimen is more likely to be effective if its composition is guided by reliable information on drug susceptibility. The design of longer regimens for MDR/RR-TB patients with additional resistance (including XDR-TB) follows a similar logic to that used for other MDR-TB patients. Ideally, all MDR-TB patients are tested for resistance to fluoroquinolones and second-line injectable agents as a minimum before starting MDR-TB treatment. Other tests for resistance to agents like bedaquiline, delamanid, linezolid, pyrazinamide and for mutation patterns commonly associated with resistance to isoniazid and the thiamides may help inform regimen choice (e.g. excluding the shorter regimen) and composition. Currently there is no approved rapid test for pyrazinamide resistance and phenotypic testing may require several weeks to produce a reliable result, implying that a decision to include or replace pyrazinamide could delay start of treatment. In many settings DST for other medicines commonly used in MDR-TB treatment is not usually reliable enough to guide regimen composition. Because of this, other elements may be necessary to determine likelihood of effectiveness (see Implementation considerations). If not already in place, the TB programme should rapidly build the capacity to undertake DST and all efforts made to ensure access to approved, rapid molecular tests. Until the capacity for second-line DST becomes available, treatment decisions may need to rely on the likelihood of resistance to medicines, based on an individual patient's clinical history and surveillance data from the country or region.

RR-TB in any patient – child or adult – in whom isoniazid resistance is absent or unknown needs to be treated with a recommended MDR-TB regimen, either a longer MDR-TB regimen to which isoniazid is added, or else a shorter MDR-TB regimen in eligible patients (see also EtD for PICO question 1). While high-dose isoniazid is not included in Table 1 above (given the rarity of its use in contemporary longer regimens for adults), this drug may still be used in patients with confirmed susceptibility or in the presence of mutations that do not usually confer complete resistance to isoniazid.

Children: The 2018 IPD MA LR was largely composed of adult patients, with only 181 of the 13,104 (1.4%) cases being under 15 years of age. Nonetheless, WHO recommendations on longer MDR-TB regimens apply to children as well as adults. Most medicines which compose longer regimens have been part of MDR-TB regimens for many years, in similar combinations, for both adults and children. Bedaquiline is approved for use in children from 6 years of age and delamanid from age 3 and over. The GDG had concerns about the feasibility of administering the correct dose to children aged 3-5 years given that the special formulation used in the trial (25 mg) was not available and only the adult tablet exists (50 mg), which is not bioequivalent and presents challenges to manipulate its contents without compromising its effectiveness. The avoidance of an injectable-containing regimen is particularly desirable in children, especially those very young and without severe disease.

Extrapulmonary TB and TB meningitis: The WHO recommendations on longer MDR-TB regimens apply also to patients with extrapulmonary disease. Adjustments may be required depending upon the specific location of disease. Treatment of MDR/RR-TB meningitis is best guided by DST of the infecting strain and by the knowledge on the properties of TB medicines to cross the blood-brain barrier. Levofloxacin and moxifloxacin penetrate well the central nervous system (CNS), as do ethionamide (or prothionamide), cycloserine (or terizidone), linezolid and imipenem. Seizures may be more common in children with meningitis treated with imipenem (meropenem is preferred for meningitis cases and children). High dose isoniazid and pyrazinamide can also reach therapeutic levels in the cerebrospinal fluid and may be useful if the strains are susceptible. PAS and ethambutol do not penetrate the CNS well and should not be counted on as effective agents for MDR-TB meningitis. Amikacin and streptomycin only penetrate the CNS in the presence of meningeal inflammation. There are little data on the CNS penetration of clofazimine, bedaguiline or delamanid.

**Pregnancy:** Am, S, Pto and Eto are usually contraindicated during pregnancy. Knowledge about the safety of Bdq and Dlm in pregnancy and while breastfeeding is sparse. It is recommended that in such cases a longer regimen be individualised to include components with a safety profile that is better established.

HIV infection: The composition of the treatment regimen for MDR-TB does not usually differ substantially for people living with HIV; a few drug-drug interactions may be avoided with careful attention (e.g. Bdq and efavirenz; see also (7)). Thioacetazone, which is no longer on the list of medicines usually recommended for use, should not be given to patients who are HIV positive or whose HIV status is unknown and HIV infection needs to be reliably excluded in the rare instances where it is being considered as part of treatment.

## Implementation considerations

The implementation of MDR-TB treatment on a large scale is feasible under programmatic conditions, as has been shown by the global expansion in the use of standardised and individualised MDR-TB regimens in low-, middle- and high-income countries worldwide, particularly in the last decade. While the current revision brings important changes to the grouping of medicines and the composition of longer MDR-TB regimens it is not expected to have a major impact on the feasibility of implementation. Changes to the regimen costs and sufficient resources to improve monitoring requirements may influence the rapidity with which the new recommendations are applied in programmes but should not stand in the way of providing good access to life-saving treatment to more patients in need. All of the agents recommended for use are available via the Global Drug Facility and most are also available in quality-assured, affordable generic formulations from other sources. Bedaquiline has been available via a donation programme for the last few years (until March 2019) and a decrease in price has been negotiated with the manufacturer for low resource settings. The latest WHO Model Lists of Essential Medicines (2017) include all agents. In August 2018 WHO and other main technical and funding partners created a *Task force to support country transition towards new recommendations for the treatment of MDR-TB* which started by developing an implementation resource in the form of Frequently Asked Questions. The Task force is expected to spearhead efforts to effect the reforms needed at country level to facilitate the uptake of the new guidance.

Where possible a patient's MDR/RR-TB strain needs to be tested for susceptibility to medicines planned for inclusion in the regimen. Access to rapid diagnostic testing which could reliably identify resistance to fluoroquinolones and injectable agents would help clinicians to decide whether the patient is eligible to the shorter MDR-TB regimen and what agents to include in a longer MDR-TB regimen (the GenoType MTBDRsl line probe assay may be used for this purpose). GenoType MTBDRsl can be used in both children and adults and as a direct and indirect test (for extrapulmonary samples). While resistance-conferring mutations to fluoroquinolones detected by the MTBDRsl assay are highly correlated with phenotypic resistance to ofloxacin, the correlation with moxifloxacin is less clear and the inclusion of moxifloxacin in a MDR-TB

regimen is best guided by phenotypic DST results.

One of the overarching outcomes of the 2018 IPD LR MA is that when a DST result indicates resistance to an agent then it is better to have it replaced. This trend also applied to medicines for which DST or the DST method used is known to be unreliable for clinical decision-making. Interpreting DST for several regimen components would present such difficulties (e.g. cycloserine, streptomycin, ethambutol). Moreover, in many settings, there may be little if any access to DST for the newer medicines for some time to come. Efforts to acquire such capacity need to be strengthened. Until DST becomes practicable for all medicines, the best possible treatment options should not be withheld and decisions have to be guided by the likelihood of resistance to these medicines, informed by the patient's clinical history and recent representative surveillance data from the area [the usual way that "likelihood of effectiveness" is assessed in the programmatic setting is on the basis of one or more of (i) confirmed susceptibility in the individual patient; (ii) confirmed susceptibility in the presumed source case; (iii) no known resistance to another drug which has cross-resistance to the medicine; (iv) rare use of the medicine in an area (possibly supported by low drug-resistance levels from surveillance activities); and (v) no previous use of the medicine in a regimen that failed to cure that same patient]. When there is uncertainty about the effectiveness of a certain agent, it can still be included in the regimen but it should be considered supernumerary to the target number of medicines needed and clinical judgment is advised to decide if the benefit from its inclusion outweighs any added toxicity, pill burden, or other downsides. The design of the regimen has to take into account the relative benefits to harms to the individual patient, including drug-drug interactions (e.g. preference for Lfx over Mfx to limit the likelihood of additive QT-interval prolongation; the inclusion of Z with Bdq on the premise that the two agen

It is expected that most patients could be treated with 4 agents at start, of which one – usually bedaquiline – would be stopped at month 6. Starting with 5 agents rather than 4 may be favoured in certain situations, namely: (i) two of the four agents may need to be stopped before the end of treatment, for instance Bdq stopped at month 6 and Lzd stopped early because of intolerance; (ii) reliable DST is not available for one or more of the agents on the regimen but background resistance to the agent is known to be high; (iii) the agents included in the regimen are unlikely to cure the patient (e.g. only a total of 2 of the agents from Group A and Group B are included in the regimen). Given that the regimen needs to have at least 3 effective agents after Bdq is stopped at 6 months, if another agent needs to be stopped because of toxicity then that medicine is replaced by another one. The replacement medicine would be chosen either from Group B (unless both Cfz and Cs/Trd are already included) or from Group C. The choice from Group C is determined by the order in which the medicines are ranked and the individual circumstances of the patient and setting.

Given the conditionality of the recommendation for the use of the shorter MDR-TB regimen, the patient and health care provider may decide for a longer treatment in patients who are otherwise eligible for the shorter MDR-TB regimen based on the individual circumstances, such as uncertainty about DST results or lack of access to second-line line probe assay (LPA); unavailability of clofazimine or another component medicine; preference for an injectable-sparing regimen; or the patient condition requires immediate start of treatment before all baseline testing can be completed. If the shorter MDR-TB regimen cannot be used the patient needs to be reassessed with a view to starting a longer MDR-TB treatment. It is important to point out that whereas a patient started on the shorter MDR-TB regimen can later be transferred to a longer regimen should the need arise, once patients are placed on a longer regimen for at least 4 weeks normally they can no longer be switched to the shorter regimen.

The GDG emphasizes the importance of patient support to complete treatment as prescribed. The high level of success achieved in both arms of the Phase III trial of delamanid points to the critical importance of ensuring medication adherence and retention to reduce treatment failure and death to a minimum. Ahead of enrolment on MDR-TB treatment, all patients should receive appropriate counselling to enable informed and participatory decision-making. Patient information material needs to reflect the new changes so that patients are appropriately informed about their treatment options. Social support to enable adherence to treatment is very important to ensure a patient-centred approach to the delivery of care.

This guideline update has concurrently revised the weight-based dosage schedules for medicines used in MDR-TB regimens in both children and adults (Annex 6 of the guidelines). The update to the dosages has benefited from the expertise of both the GDG members as well as a very extensive consultation of other specialists in different fields and was based on the latest knowledge available for the optimal use of the medicines involved. Adherence to the schedules is advised as far as possible. Manipulation of tablets (splitting, crushing, dissolving in water) beyond their indications is to be reduced to the minimum possible given that it will interfere in their bioavailability. This is particularly problematic with the delamanid tablet preparation.

## **Monitoring and evaluation**

Patients on longer MDR-TB treatment regimens need to be monitored for response to treatment and for safety using reasonable schedules of appropriate clinical and laboratory testing. Electrocardiography may be indicated given that more regimens in future are expected to have two or three agents with QT-interval prolonging properties given concurrently.

A separate recommendation on the use of culture and microscopy to monitor bacteriological response during treatment is being made in the 2018 update of the guidelines (see PICO question 7).

Standardized approaches to the surveillance of drug-resistance through continuous monitoring of diagnostic DST (including the use of sequencing(10)) and for the assignment of treatment outcomes in annual patient cohorts have been available in WHO normative documents since many years. The systematic monitoring of adverse events during and after the end of treatment is a recent introduction in TB programmes and experience in their implementation is still developing in many countries. Its rationale is largely defined by frequent use of new and re-purposed medications in MDR-TB treatment regimens in the world, at times in combinations for which there has been very limited experience of use. Guidance on monitoring safety is provided in *Active Tuberculosis Drug-Safety Monitoring (aDSM): Framework for Implementation* and the *Companion Handbook to the WHO Guidelines for the Programmatic Management of Drug Resistant Tuberculosis*(11),(12).

# Research priorities

In addition to summarizing the available evidence, the reviews undertaken for this update revealed a number of gaps in current knowledge about critical areas for the treatment for MDR/RR-TB. The estimates of effect were commonly assigned low or very low certainty ratings. The quality of evidence was one of the main reasons why most of the recommendations made in these guidelines are conditional. Some gaps persist from the ones identified in previous TB treatment guidelines.

Implementation research, studies of resource use and inclusion of indicators of quality of life would be relevant to much of the priority questions in the programmatic management of drug-resistant tuberculosis. When completing the Evidence to Decision tables for PICO questions 2 and 3 (and others) the GDG members did not have at their disposal any studies focused on how patients, caregivers and other stakeholders value different outcomes such as cure, treatment failure and relapse, death and serious adverse events and therefore the GDG expressed their own judgement on behalf of end users of the guidelines. Moreover, the following areas for prioritization of research are highlighted by the GDG:

- The optimal combination of medicines and approach towards regimen-design for adults and children with MDR/RR-TB with or without additional resistance to key agents
- Randomized controlled trials, especially involving the new drugs and regimens, remain rare. The release of results from the first Phase III trials for MDR-TB has led to substantial debate about the clinical relevance of the design and endpoints chosen for these studies, requiring at times additional, off protocol exploration of data for potential added value.
- Inclusion and separate reporting of outcomes for key subgroups, especially children and HIV-positive individuals on treatment, in randomized controlled studies.
- Pharmacokinetic studies to determine optimal drug dosing and safety (especially in pregnancy).
- Complete recording of adverse events and standardized data recording on organ class, seriousness, severity, and certainty of association, to allow meaningful comparison of the association between adverse events and exposure to different medicines between studies, patient subgroups, and different regimens.
- Determination of the minimum number of drugs and treatment duration (especially in patients previously treated for MDR-TB).
- Improved diagnostics and drug-susceptibility testing methods, especially for medicines for which no rapid molecular methods are currently available in the field.
- Further research and developments that would be particularly helpful for the following agents:
- Levofloxacin: optimization of the dose (the Opti-Q study will provide new information on this even by late 2018)
- Bedaquiline: use in children to determine optimal pharmacokinetic properties; revised cost effectiveness analyses based on the IPD meta-analysis.
- Linezolid: optimization of the dose and duration in both adults and children; patient predictors for adverse reactions
- Clofazimine: optimization of dose especially in children; any added value in use of a loading dose; availability of DST methods
- Cycloserine / Terizidone: differences in efficacy between the two medicines; approaches to test for susceptibility to them; best practices in psychiatric care for persons on these medicines
- Delamanid: better understanding of its role in MDR-TB regimens, including in children (PK/PD), PLHIV and pregnant women; mechanisms of development of drug resistance
- Pyrazinamide: molecular testing for resistance (pursuing either line probe assay or another approach)
- Carbapenems: the potential role and cost-effectiveness of ertapenem (that can be given intramuscularly) as a substitute for Mpm and Imp-Cln
- Amikacin: the safety and effectiveness of three-times weekly administration at a higher dose (about 25mg/kg/day) (13)

#### **REFERENCES SUMMARY**

- 1. Ahuja SD, Ashkin D, Avendano M, Banerjee R, Bauer M, Bayona JN, et al. Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: an individual patient data meta-analysis of 9,153 patients. PLoS Med. 2012;9(8):e1001300.
- 2. Harausz EP, Garcia-Prats AJ, Law S, Schaaf HS, Kredo T, Seddon JA, et al. Treatment and outcomes in children with multidrug-resistant tuberculosis: A systematic review and individual patient data meta-analysis. Metcalfe JZ, editor. PLOS Med. 2018 Jul 11;15(7):e1002591.
- 3. Seddon JA, Hesseling AC, Godfrey-Faussett P, Schaaf HS. High treatment success in children treated for multidrug-resistant tuberculosis: an observational cohort study. Thorax. 2014 May 1;69(5):458–64.
- 4. Ahmad N, Ahuja SD, Akkerman OW, Alffenaar J-WC, Anderson LF, Baghaei P, et al. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. The Lancet. 2018 Sep;392(10150):821–34.
- 5. WHO position statement on the use of delamanid for multidrug-resistant tuberculosis (WHO/CDS/TB/2018.1) [Internet]. Geneva, World Health Organization; 2018. Available from: http://www.who.int/tb/publications/2018/WHOPositionStatementDelamanidUse.pdf
- 6. The use of delamanid in the treatment of multidrug-resistant tuberculosis in children and adolescents: interim policy guidance (WHO/HTM/TB/2016.14) [Internet]. Geneva, World Health Organization. 2016. Available from: http://apps.who.int/iris/bitstream/10665/250614/1/9789241549899-eng.pdf
- 7. University of Liverpool. HIV Drug Interactions [Internet]. Available from: https://www.hiv-druginteractions.org/checker
- 8. Ibrahim M, Andries K, Lounis N, Chauffour A, Truffot-Pernot C, Jarlier V, et al. Synergistic Activity of R207910 Combined with Pyrazinamide against Murine Tuberculosis. Antimicrob Agents Chemother. 2007 Mar 1;51(3):1011–5.
- 9. Veziris N, Ibrahim M, Lounis N, Chauffour A, Truffot-Pernot C, Andries K, et al. A Once-Weekly R207910-containing Regimen Exceeds Activity of the Standard Daily Regimen in Murine Tuberculosis. Am J Respir Crit Care Med. 2009 Jan;179(1):75–9.
- 10. The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide (WHO/CDS/TB/2018.19). [Internet]. Geneva, World Health Organization / FIND; 2018. Available from: http://apps.who.int/iris/bitstream/handle/10665/274443/WHO-CDS-TB-2018.19-eng.pdf
- 11. Active tuberculosis drug-safety monitoring and management (aDSM). Framework for implementation (WHO/HTM/TB/2015.28) [Internet]. Geneva, World Health Organization; 2015. Available from: http://apps.who.int/iris/bitstream/10665/204465/1/WHO HTM TB 2015.28 eng.pdf
- 12. Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis. (WHO/HTM/TB/2014.11) [Internet]. Geneva, World Health Organization; 2015. Available from: http://apps.who.int/iris/bitstream/10665/130918/1/9789241548809 eng.pdf
- 13. Technical report on the pharmacokinetics and pharmacodynamics (PK/PD) of medicines used in the treatment of drug-resistant tuberculosis (WHO/CDS/TB/2018.6) [Internet]. Geneva, World Health Organization; 2018. Available from: http://apps.who.int/iris/bitstream/10665/260440/1/WHO-CDS-TB-2018.6-eng.pdf

# PICO QUESTION 4

In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines are outcomes safely improved with an intensive phase shorter or longer than eight months?

POPULATION:	Patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens conforming to WHO guidelines				
INTERVENTION: An intensive phase of 8 months' duration					
COMPARISON: An intensive phase of less than 8 months' duration (5-5.99 months vs. 7-8.49 months on an injectable agent)					
MAIN OUTCOMES:	Treatment failure or relapse versus treatment success				
SETTING:	Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in low-, middle- and high-income settings, using hospital or ambulatory models of care.				
PERSPECTIVE:	The findings of these analyses are expected to influence the continued validity in its present form of the conditional recommendation made by WHO in 2011 on a duration of about 8 months for the intensive phase for patients on longer MDR-TB regimens who had not been previously treated and it was recommended that this duration be modified according to the patient's response to therapy. Furthermore, the total duration of treatment is addressed in PICO question 5 while PICO question 6 addresses the minimum duration of treatment after culture conversion.				
BACKGROUND:	The recommendation in the <i>Guidelines for the Programmatic Management of Drug Resistant Tuberculosis, 2011 Update</i> was based on an individual patient data meta-analysis inclusive of data from 32 observational studies with over 9,000 patient records (1),(2). This analysis showed an association between treatment success and an intensive phase lasting between 7.1 and 8.5 months. Given significant changes introduced to the longer regimens used worldwide over the last decade, with the global expansion of the use of later-generation fluoroquinolones, bedaquiline and linezolid, it was deemed timely to review evidence for this question to determine if 8 months remains optimal.  The analysis to inform PICO question 4 looked at treatment failure and relapse versus success for different durations of the intensive phase. The individual patient data meta-analysis for the 2018 update included records from 3,750 patients for the primary analyses, from 42 observational studies; of whom 2,720 were treated with an individualized MDR-TB regimen and 1,030 were treated with standardized MDR-TB regimens. These 3,750 patients were a sub-set of the individual patient dataset used for PICO question 2, which comprised 13,104 patients overall, from 53 studies in 40 countries. To address PICO question 4, 9,354 records were excluded from the larger individual patient dataset for the following reasons (lost to follow up: n=2,261; died: n=2,043; did not receive an injectable: n=1,094; no information on duration of injectable: n=2,341; number of medicines likely to be effective less than four plus				
	pyrazinamide: n=1,450; duration of injectable greater than 20 months: n=165 (GRADE summary of evidence table in Annex 8)				

# **ASSESSMENT**

Problem  Is the problem a priority?								
JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS						
<ul> <li>No</li> <li>Probably no</li> <li>Probably yes</li> <li>Yes</li> <li>Varies</li> <li>Don't know</li> </ul>	About 558,000 new cases of rifampicin- and multidrug-resistant TB (MDR/RR-TB) cases are estimated to emerge worldwide each year and 230,000 cases die. Globally only about one fourth of newly emergent MDR-TB patients have been reported to start a second-line TB treatment annually in recent years.  Outcomes of MDR-TB treatment on a global level are poor with much loss to follow up and death, and only 55% of cases reported to have a successful outcome at the end of treatment. The need for better treatment for these patients is thus a priority in an effort to save lives and reduce transmission and chronicity.	MDR-TB is a global challenge and access to treatment often problematic, with regimens typically being long, toxic and expensive.  There is clearly an interest in reducing the duration of treatment, simplifying the administration of the treatment regimen, and providing patients with a safer combination of medicines that can cure the large majority of cases.						

### **Desirable Effects**

How substantial are the desirable anticipated effects?

#### JUDGEMENT

#### RESEARCH EVIDENCE

#### o Trivial

- Small
- o Moderate
- o Large
- o Varies
- o Don't know

# PRIMARY ANALYSIS: DURATION OF THE INTENSIVE PHASE OF TREATMENT (ALL REGIMENS)<sup>1</sup>

	FAIL/RELAPSE vs. SUCCESS - PS Score Matched Analysis							
Injectable Duration	Events/Total	N Pairs	<u>aOR</u> (95% CI)	aRD (95% CI)				
1.00-4.99 months	29/416	416	0.4 (0.1, 1.8)	-3 (-9, 3)				
5.00-5.99 months	145/730	623	0.9 (0.4, 2.3)	0 (-3, 3)				
6.00-6.99 months	24/554	554	0.2 (0.0, 1.1)	-3 (-7, 0.4)				
7.00-8.49 months	38/623	-	1.0 (reference)	0 (reference)				
8.50-11.99 months	54/737	623	0.9 (0.4, 1.8)	0 (-4, 3)				
12.00-19.99 months	110/690	623	3.6 (1.1, 11.3)	5 (1, 9)				

1. The analysis excludes individuals based on fewer than 5 effective drugs or 4+PZA. Excludes individuals on requirement of intensive phase of treatment duration information. All regimens considered.

Outcomes	Nº of participants (studies)	of the effect evidence (95%		Anticipated absolute effects* (95% CI)		
	Follow up	(GRADE)	CI)	Risk with an intensive phase of less than 8 months	Risk difference with an intensive phase of 8 months	
Treatment Failure/Relapse vs. Treatment Success: 5-	nonths on an injectable studies) (42 LOW LOW		<b>aOR 0.9</b> (0.4 to	Study population		
5.99 months on an injectable vs. 7-8.49 months on an injectable		2.3)	20 per 100	2 fewer per 100 (12 fewer to 26 more)		
Treatment Failure/Relapse vs. Treatment Success :6-	ment Success :6- other on an injectable observational studies)  (42 observational studies)  (0.0 to 1.1)			Study population		
6.99 months on an injectable vs. 7-8.49 months on an injectable		4 per 100	3 fewer per 100 (4 fewer to 0 fewer)			

#### ADDITIONAL CONSIDERATIONS

When comparing the aOR for treatment failure or relapse vs. success the minimum value when compared to an intensive phase of 7-8.5 months was 6-7 months. This effect was also observed when the analysis was restricted to patients who received at least a later generation fluoroquinolone (if sensitive), or bedaquiline, or linezolid, or clofazimine, or a carbapenem. A similar pattern was observed when the regimen had only 4 effective agents (or 3 plus pyrazinamide). The GDG judged the absolute effects as small overall. In XDR-TB patients a duration of 5-5.99 months was associated with lower risk of failure although number of observations was very small.

### **Undesirable Effects**

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE					ADDITIONAL CONSIDERATIONS	
o Large o Moderate o Small							The risk of the treatment failure or relapse was slightly different when comparing a 6-7 month intensive phase and one of 7-8.5
o Trivial ● Varies o Don't know	participants (studies)	participants	Certainty of the evidence	effect	(95% CI)		month duration, with a larger confidence interval for a 6 month intensive phase. The GDG also acknowledged that a shorter duration of the injectable phase would be preferred by most patients and is beneficial for the health services too (information on adverse events were not available to the GDG to make a definitive judgement about this).
		(GRADE)	ČI)	Risk with an intensive phase of less than 8 months	Risk difference with an intensive phase of 8 months		
	· · · · · · · · · · · · · · · · · · ·	1353 (42	⊕⊕○○ LOW	<b>aOR 0.9</b> (0.4 to	Study population		
	5.99 months on an injectable vs. 7-8.49 months on an injectable	observational studies)	LOW	2.3)	20 per 100	2 fewer per 100 (12 fewer to 26 more)	
	Treatment Failure/Relapse vs. Treatment Success :6-	, , , , , , , , , , , , , , , , , , ,	LOW (0.0 to	OW (0.0 to	<b>/</b>		
	6.99 months on an injectable vs. 7-8.49 months on an injectable	observational studies)			4 per 100	3 fewer per 100 (4 fewer to 0 fewer)	

# Certainty of evidence What is the overall certainty of the evidence of effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low  ■ Low  o Moderate  o High  o No included studies	These data were derived from an individual patient data meta analysis which comprised a total of 3750 patients for the primary analyses, from 42 observational studies; of whom 2720 were treated with an individualized MDR-TB regimen and 1030 were treated with a standardized MDR-TB regimen. These 3750 patients were a sub-set of the individual patient dataset used for PICO 2, which comprised 13,104 patients overall, from 53 studies and 40 countries. The overall certainty of the evidence was classified as low.	The certainty of the evidence is low based on the fact that the majority of studies included in the IPD meta-analysis are observational. The analyses of the 2018 IPD the data were matched (exactly) for resistance to a fluoroquinolone, resistance to a second line injectable agent and income level (according to the World Bank Atlas method). In addition, propensity score matching was used for a number of other important covariates. Patients who receive longer therapy may be those who are sicker, which may introduce bias that is difficult to control for.

# Values

Is there important uncertainty about or variability in how much people value the main outcomes?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Important uncertainty or variability o Possibly important uncertainty or variability • Probably no important uncertainty or variability o No important uncertainty or variability	No research evidence was identified.	Although no research evidence was identified, the GDG concluded that there was probably no important uncertainty of variability in how much people would value a favourable treatment outcome, given an appropriate duration of the intensive phase. A high value might be placed on a favourable treatment outcome by both patients and health care workers and many patients may also place a high value on avoiding a long course of treatment (or a longer intensive phase which includes an injectable agent) due to the burden and inconvenience of treatment and the pain associated with injections.

# **Balance of effects**

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Favors the comparison o Probably favors the comparison o Does not favor either the intervention or the comparison ● Probably favors the intervention o Favors the intervention o Varies o Don't know		The desirable effects were described as small, while the undesirable effects were described as being varied. Therefore, the GDG concluded that overall the balance of effects probably favours the intervention.

# Resources required

How large are the resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Large costs o Moderate costs o Negligible costs and savings o Moderate savings o Large savings o Varies • Don't know	No research evidence was identified.	Although no research evidence was identified, the GDG thought that a longer intensive phase inclusive of an injectable agent might require more resources including staffing, medication, directly observed therapy and monitoring (e.g. audiometry, kidney function). Costs are also incurred by the patient for each month that a visit to a health facility is required, which may be particularly important given the need for regular intramuscular injections during the intensive phase.

# Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low o Low o Moderate o High • No included studies	No research evidence was identified.	

# **Cost effectiveness**

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
O Favors the comparison O Probably favors the comparison O Does not favor either the intervention or the comparison O Probably favors the intervention O Favors the intervention O Varies  No included studies	No research evidence was identified.	

**Equity**What would be the impact on health equity?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Reduced o Probably reduced o Probably no impact o Probably increased o Increased ● Varies o Don't know		Although no research evidence was identified, the GDG concluded that the impact on health equity might vary.

Acceptability
Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
<ul> <li>○ No</li> <li>○ Probably no</li> <li>● Probably yes</li> <li>○ Yes</li> <li>○ Varies</li> <li>○ Don't know</li> </ul>	No research evidence was identified.	Although no research evidence was identified, the GDG thought that a shorter intensive phase would likely be more acceptable to patients and health care workers as this would shorten the amount of time needed for the administration of injections (and other oral medicines). It is expected that the need for an injectable agent in the longer regimen will diminish
		with time as more patients receive oral- only treatment

# Feasibility

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o No o Probably no ● Probably yes o Yes o Varies O Don't know	No research evidence was identified.	The intervention of a 6-7 month intensive phase for patients with MDR/RR-TB was thought to be feasible to implement as the current recommendation for 8 months is being widely sued. Therefore, shortening the period should be relatively feasible to
		implement.

# **SUMMARY OF JUDGEMENTS**

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	Yes		Varies	Don't know
DESIRABLE EFFECTS	Trivial	Small	Moderate	Large		Varies	Don't know
UNDESIRABLE EFFECTS	Large	Moderate	Small	Trivial		Varies	Don't know
CERTAINTY OF EVIDENCE	Very low	Low	Moderate	High			No included studies
VALUES	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability			
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention Favors the intervention		Varies	Don't know
RESOURCES REQUIRED	Large costs	Moderate costs	Negligible costs and savings	Moderate savings Large savings		Varies	Don't know
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	Low	Moderate	High			No included studies
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	No included studies
EQUITY	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know
ACCEPTABILITY	No	Probably no	Probably yes	Yes		Varies	Don't know
FEASIBILITY	No	Probably no	Probably yes	Yes		Varies	Don't know

#### TYPE OF RECOMMENDATION

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	Strong recommendation for the intervention
0	0	0	•	0

#### CONCLUSIONS

#### Recommendation

In MDR/RR-TB patients on longer regimens containing an injectable agent, an intensive phase of 6-7 months is suggested for most patients, and the duration may be reduced or increased according to the patient's response to therapy (conditional recommendation, low certainty in the estimates of effect).

#### **Justification**

The duration of the intensive phase in the treatment of MDR-TB patients with longer second-line regimens has been associated with likelihood of treatment success and failure (2). The evidence that informed this recommendation was a subset of the data from the 2018 IPD-MA (see Background section on first page). This relied heavily on observational studies with few studies conducted under randomized controlled conditions. One of the main concerns is that, particularly in patients on individualized regimens, treatment may have been prolonged if a patient was not improving clinically, and so patients who received longer therapy may be those who were sicker. Attempts to control for bias and confounding in the evidence review are unlikely to have adjusted for all factors. As a result, the quality of the evidence was judged to be low.

# **Subgroup considerations**

**Children:** This recommendation applies to both children and adults. However, use of amikacin or streptomycin in children should only be resorted to when other options are not possible, when testing confirms susceptibility and the possibility to monitor for ototoxicity and nephrotoxicity is present.

Pregnant women: due to the potential for teratogenic effects, injectable agents are usually contraindicated in pregnancy and therefore this recommendation will be of very limited relevance.

MDR/RR-TB alone or with additional resistance: The analysis indicated that XDR-TB patients could benefit from an intensive phase of 5-6 months although the number of cases was small and the potential contribution of injectable agents when the strain is resistant is doubtful. A regimen without an aminoglycoside or an intensive phase and including newer medicines would be advisable in this subgroup.

# Implementation considerations

This recommendation applies only to patients in whom amikacin or streptomycin is being used as the evidence underpinning it included only patients who were receiving an injectable agent during the intensive phase. For patients on an all oral regimen for MDR/RR-TB, the duration of treatment is determined by the total duration of treatment and the duration of treatment post culture conversion (see PICO questions 5 and 6) OR the usual recommended duration of use of certain medicines such as bedaquiline and delamanid. It is important to point out that the initial period of treatment during which bedaquiline or delamanid are usually given does not equate with the intensive phase unless an injectable agent is used concurrently.

National TB programmes may find it more practical and straightforward to apply a fixed duration of intensive phase (e.g. 6 months) to facilitate implementation. The clinician may find it necessary to prolong the intensive phase if there are grounds for doing so (e.g. prolonged positivity of sputum). In the case of the emergence of toxicity associated with the injectable agent a change of the regimen becomes necessary and the continuation phase started with the revised treatment.

## Monitoring and evaluation

Patients on longer MDR-TB treatment regimens need to be monitored for response to treatment, treatment failure and for safety using reasonable schedules of relevant clinical and laboratory testing (3),(4). Monitoring response to treatment is done through regular history taking, physical examination, chest radiograph and laboratory monitoring. Conversion of bacteriological status using smear microscopy or culture is an important means of assessing response and most patients are usually expected to have converted to negative within the first few months of starting treatment. Persistence of culture positivity beyond that point and approaching the end of the intensive phase is a trigger for a review of the regimen and the performance of DST. Frameworks for the surveillance of bacteriological status, drug-resistance and the assignment of outcomes have been fairly standardized in past years (5). In contrast, the systematic monitoring of adverse events during and after the end of treatment needs to be strengthened in most TB programmes given the relative novelty of active pharmacovigilance within national TB programmes. In the case of this recommendation it is important to monitor for hearing loss and kidney function given the use of the injectable agent. The rationale for aDSM is largely supported by the increasing use worldwide of combinations of new and re-purposed medications in MDR-TB treatment regimens. The toxicity of certain agents (such as linezolid) is increased with duration of use and may limit their continued employment in a patient, and at times the complete cessation of treatment. The prospective collection of accurate data for key variables at case-based level using an electronic register is strongly advised in the best interests of the individual patient and to inform local and global policy revisions (6).

### **Research priorities**

In the process of summarizing the available evidence, the reviews undertaken for this update revealed a number of gaps in current knowledge about critical areas for the treatment for MDR/RR-TB.

Amongst these, the GDG remarked that further research is required on adverse events during treatment of patients with MDR/ RR-TB, including in the intensive phase, preferably with standardized data reporting on the seriousness and severity of the adverse event, organ class affected, and certainty of association. It would be helpful to collect safety and effectiveness data for patients treated with three-times weekly amikacin given at a higher dose (about 25mg/kg/day) (7).

#### **REFERENCES SUMMARY**

- 1. Guidelines for the programmatic management of drug-resistant tuberculosis, 2011 Update. (WHO/HTM/TB/2011.6) [Internet]. Geneva, World Health Organization. 2011. Available from: http://whqlibdoc.who.int/publications/2011/9789241501583 eng.pdf
- 2. Ahuja SD, Ashkin D, Avendano M, Banerjee R, Bauer M, Bayona JN, et al. Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: an individual patient data meta-analysis of 9,153 patients. PLoS Med. 2012;9(8):e1001300.
- 3. Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis. (WHO/HTM/TB/2014.11) [Internet]. Geneva, World Health Organization; 2015. Available from: http://apps.who.int/iris/bitstream/10665/130918/1/9789241548809 eng.pdf
- 4. Active tuberculosis drug-safety monitoring and management (aDSM). Framework for implementation.(WHO/HTM/TB/2015.28) [Internet]. Geneva, World Health Organization; 2015. Available from: http://apps.who.int/iris/bitstream/10665/204465/1/WHO HTM TB 2015.28 eng.pdf
- 5. Definitions and reporting framework for tuberculosis 2013 revision (WHO/HTM/TB/2013.2) [Internet]. Geneva, World Health Organization; 2013. Available from: http://apps.who.int/iris/bitstream/10665/79199/1/9789241505345 eng.pdf
- 6. Electronic recording and reporting for tuberculosis care and control (WHO/HTM/TB/2011.22) [Internet]. Geneva, World Health Organization; 2012. Available from: http://whqlibdoc.who.int/publications/2012/9789241564465 eng.pdf
- 7. Technical report on the pharmacokinetics and pharmacodynamics (PK/PD) of medicines used in the treatment of drug-resistant tuberculosis (WHO/CDS/TB/2018.6) [Internet]. Geneva, World Health Organization; 2018. Available from: http://apps.who.int/iris/bitstream/10665/260440/1/WHO-CDS-TB-2018.6-eng.pdf

#### **PICO QUESTION 5**

**PERSPECTIVE:** 

**BACKGROUND:** 

In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines are outcomes safely improved with a total duration shorter or longer than twenty months?

POPULATION: Patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens conforming to WHO guidelines

INTERVENTION: A total treatment duration of 18 to 20 months

COMPARISON: A total treatment duration lasting less than 17.50 months or more than 19.99 months

MAIN OUTCOMES: Treatment failure or relapse versus treatment success

SETTING: Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in low-, middle- and high-income settings, using hospital or ambulatory models of care.

The findings of these analyses are expected to influence the continued validity in its present form of the conditional recommendation made by WHO in 2011 for a total duration of 20 months in the treatment of newly diagnosed MDR-TB. Furthermore, the duration of the intensive phase is addressed in PICO 4, while the minimum duration of treatment post culture

conversion is addressed in PICO 6.

The recommendation in the *Guidelines for the Programmatic Management of Drug Resistant Tuberculosis, 2011 Update* was based on an individual patient data meta-analysis inclusive of data from 32 observational studies with over 9,000 patient records (1),(2). In 2011, for the total treatment duration, the adjusted relative risk for cure peaked between 18.6 and 21.5 months for patients who had no previous MDR-TB treatment. The peak occurred later in patients who had been previously treated for MDR-TB (27.6–30.5 months) but there was a less clear incremental trend and less observations than in the case of patients who had not been previously treated for MDR-TB. Given recent developments in regimen composition including the more widespread use of moxifloxacin, bedaquiline, linezolid and longer intensive phase, the question on the optimal duration of treatment for patients for MDR/RR-TB became topical again. One PICO question was thus devoted to the total duration of longer MDR-TB treatment regimens.

For the 2018 update, the data were derived from a fresh individual patient data meta-analysis with 6,356 patients from 51 observational studies for the primary analysis. These patients were from a sub-set of the individual patient dataset used for PICO 2 (comprised of 13,104 patients overall from 53 studies in 40 countries). Of 6,356 patients, 5,352 were treated with an individualized MDR-TB regimen and 1,004 were treated with a standardized MDR-TB regimen. To address PICO question 5, 6,748 records were excluded from the larger individual patient dataset for the following reasons (primary exclusions: lost to follow up: n=2261; died: n=2043; treatment duration not available: n=230; number of effective drugs less than five or less than four plus pyrazinamide: n=2072; treatment duration less than six months: n=52; treatment duration greater than or equal to 36 months: n=90). Information on relapse was only available for 388 of the included patients (with 9 relapse events). (GRADE summary of evidence table in Annex 8).

#### **ASSESSMENT**

Problem Is the problem a priority?					
JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS			
o No o Probably no o Probably yes ● Yes o Varies o Don't know	About 558,000 new cases of rifampicin- and multidrug-resistant TB (MDR/RR-TB) cases are estimated to emerge worldwide each year and 230,000 cases die. Globally only about one fourth of newly emergent MDR-TB patients have been reported to start a second-line TB treatment annually in recent years.  Outcomes of MDR-TB treatment on a global level are poor with much loss to follow up and death, and 55% of cases reported to have a successful outcome at the end of treatment. The need for better treatment for these patients is thus a priority in an effort to save lives and reduce transmission and chronicity.	MDR-TB is a global challenge and access to treatment often problematic, with regimens typically being long, toxic and expensive.  There is clearly an interest in reducing the duration of treatment, simplifying the administration of the treatment regimen, and providing patients with a safer combination of medicines that can cure the large majority of cases.			

### **Desirable Effects**

How substantial are the desirable anticipated effects?

#### JUDGEMENT RESEARCH EVIDENCE ADDITIONAL CONSIDERATIONS PRIMARY ANALYSIS: TOTAL REPORTED DURATION OF TREATMENT (ALL REGIMENS)1 o Trivial The reference range of 18 to 20 months was favoured by Small the GDG. The desirable effects (including treatment Moderate success) may vary by duration of treatment. The GDG FAIL/RELAPSE vs. SUCCESS, PS Matched Analysis **Total Treatment Duration** o Large aRD acknowledged that there may be some confounding by $I^2$ Events/Total N Pairs (95% CI) (95% CI) (Months) o Varies indication as patients who are improving clinically may 6.00-11.99 67/100 100 78.0 (6.5, 936.1) 63 (52, 74) o Don't know have their treatment stopped at the recommended time 12.00-17.49 73/418 418 7.2 (3.3, 15.5)\* 15 (10, 19) (i.e. patients who receive longer therapy may be those 17.50-19.99 31/1376 0 (reference) 1.0 (reference) 20.00-21.99 87/1407 1376 2.1 (0.7, 6.1) 2 (-0.3, 3) who are sicker and who have not improved clinically). 22.00-23.99 166/1436 1376 4 (2.6, 6.4)\* 6 (4, 9) 24.00-25.99 79/1067 1067 2.9 (1.6, 5.4)\* 5 (3, 7) 26.00-35.99 89/552 552 5.9 (2.5, 14.1)\* 13 (9, 17) 1. Primary Analysis - excludes individuals based on fewer than 5 effective drugs or 4+PZA. Excludes individuals on requirement of treatment duration information. All regimens included. \* Fixed effects analyses used to estimate aOR and 95% CI due to non-convergence of the random effects model (small numbers in one group)

### **Undesirable Effects**

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE		ADDITIONAL CONSIDERATIONS				
o Large o Moderate o Small o Trivial o Varies • Don't know	Total Treatment Duration (Months)   6.00-11.99   12.00-17.49   17.50-19.99   22.00-23.99   24.00-25.99   26.00-35.99   1. Primary Analysis – exclud requirement of treatment d	FAIL/RELAI Events/Total 67/100 73/418 31/1376 87/1407 166/1436 79/1067 89/552 es individuals bas	PSE vs. SUCCESS, PS  N Pairs  100  418  - 1376  1376  1067  552  sed on fewer than 5 effected	Matched Analysis  aOR (95% CI)  78.0 (6.5, 936.1)  7.2 (3.3, 15.5)*  1.0 (reference)  2.1 (0.7, 6.1)  4 (2.6, 6.4)*  2.9 (1.6, 5.4)*  5.9 (2.5, 14.1)*  etive drugs or 4+PZA. Exe	aRD (95% CI) 63 (52, 74) 15 (10, 19) 0 (reference) 2 (-0.3, 3) 6 (4, 9) 5 (3, 7) 13 (9, 17)	I <sup>2</sup>	ADDITIONAL CONSIDERATIONS  There may be significant harms associated with further shortening of the regimen below 17.5 months (an increase in the outcomes of failure/ relapse is noted for patients receiving less than 17.5 months of treatment). The GDG also acknowledged that when patients are treated for shorter periods they may experience fewer adverse events (although information on adverse events were not available).
	<ul> <li>Fixed effects analyses used to esti group)</li> </ul>	mate aOR and 95	% CI due to non-conver	rgence of the random effec	ts model (small numbe	ers in one	

# **Certainty of evidence**

What is the overall certainty of the evidence of effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
● Very low ○ Low ○ Moderate ○ High ○ No included studies	The evidence is generated from observational studies included in the 2018 IPD meta-analysis - inclusive of 6356 patients from 51 studies; of whom 5325 were treated with an individualized MDR-TB regimen and 1004 were treated with a standardized MDR-TB regimen. In addition to the nature of the studies from which these data were generated, the certainty in the evidence was further downgraded to very low due to the width of the confidence interval (i.e. rated downwards for imprecision).	The certainty of the evidence is very low based on the fact that the majority of studies included in the IPD-MA are observational. In addition, there were relatively few patients who had information on relapse (n=388, with 9 reported relapses). The analyses of the 2018 IPD data were matched (exactly) for resistance to a fluoroquinolone, resistance to a second line injectable agent and income level (according to the World Bank Atlas method). In addition, propensity score matching was used for a number of other important covariates. Therefore, efforts were made to adjust for potential confounders, however residual confounding is likely.

# Values

Is there important uncertainty about or variability in how much people value the main outcomes?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
O Important uncertainty or variability O Possibly important uncertainty or variability O Probably no important uncertainty or variability  No important uncertainty or variability	No research evidence was identified.	Although no research evidence was identified, the GDG considered that there was no important uncertainty or variability in how much people would value the outcomes.

# **Balance of effects**

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
O Favors the comparison O Probably favors the comparison O Does not favor either the intervention or the comparison  Probably favors the intervention O Favors the intervention O Varies O Don't know	The desirable effects included a lowered risk of treatment failure and relapse in a regimen of a total duration of 18-20 months. The undesirable effects were unknown as information on harms was not available, however the GDG acknowledged that with a longer duration of treatment the likelihood of harms may be greater and the likelihood of treatment interruption as well. Therefore, the GDG felt that the balance of effects probably favoured the intervention.	The reference range of 18 to 20 months was favoured by the GDG. The GDG acknowledged that the magnitude of adverse events is unknown, but they may be acceptable (or, at least insufficient to lead to discontinuation of that drug), given the observed benefits of improved treatment outcomes.

# **Resources required**

How large are the resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Large costs o Moderate costs o Negligible costs and savings ● Moderate savings o Large savings o Varies o Don't know	No research evidence was identified.	Although no research evidence was identified, the GDG acknowledged that a longer treatment duration will require more resources including staffing for patient monitoring, hospitalization, medication, testing etc. Costs are also incurred by the patient for each month that the treatment is prolonged. Moderate savings may be made considering the reduction of the MDR/RR-TB regimen from 20 months (previous recommendation) duration to 18 to 20 months (current recommendation).

# **Certainty of evidence of required resources** What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
• Very low o Low o Moderate	No research evidence was identified.	Although no research evidence was identified, the GDG acknowledged that the certainty of the evidence of required resources was very low.
<ul><li>O High</li><li>O No included studies</li></ul>		

# **Cost effectiveness**

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
<ul> <li>o Favors the comparison</li> <li>o Probably favors the comparison</li> <li>o Does not favor either the intervention or the comparison</li> <li>o Probably favors the intervention</li> <li>o Favors the intervention</li> <li>o Varies</li> <li>No included studies</li> </ul>	No research evidence was identified.	

_		
- 61	$\Pi\Pi T$	V
	шь	7

What would be the impact on health equity?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Reduced	No research evidence was identified.	Although no research evidence was identified on equity in
o Probably reduced		particular, the evidence assessed suggest that a total
o Probably no impact		regimen duration of 18-20 months is effective, in practice
<ul> <li>Probably increased</li> </ul>		being comparable to the previous recommendation of 20
o Increased		months. The GDG therefore felt that the impact on health
o Varies		equity might even increase if more patients are treated
o Don't know		with this duration, including retreated MDR-TB patients.

Acceptability
Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o No o Probably no ● Probably yes o Yes o Varies o Don't know	No research evidence was identified.	Although no research evidence was identified, the GDG considered that a longer regimen that is comparable and even shorter than the one recommended since 2011 may be more acceptable to patients if cure is as likely. Health care staff may find it less acceptable if they have concerns that a regimen of this duration may be less effective. Concomitant changes to the regimen composition (e.g. preference for an oral-only regimen, inclusion of more effective agents) are expected to increase the overall acceptability of the new longer MDR-TB regimens.

**Feasibility**Is the intervention feasible to implement?

JUDGEMENT RESI	SEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o No o Probably no o Probably yes  ● Yes o Varies o Don't know		Although no research evidence was identified, the GDG acknowledged that the current recommendation is similar to, or shorter than the current recommendation, and therefore it would be feasible to implement it.

# **SUMMARY OF JUDGEMENTS**

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	Yes		Varies	Don't know
DESIRABLE EFFECTS	Trivial	Small	Moderate	Large		Varies	Don't know
UNDESIRABLE EFFECTS	Large	Moderate	Small	Trivial		Varies	Don't know
CERTAINTY OF EVIDENCE	Very low	Low	Moderate	High			No included studies
VALUES	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability			
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	Don't know
RESOURCES REQUIRED	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	Low	Moderate	High			No included studies
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	No included studies
EQUITY	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know
ACCEPTABILITY	No	Probably no	Probably yes	Yes		Varies	Don't know
FEASIBILITY	No	Probably no	Probably yes	Yes		Varies	Don't know

#### TYPE OF RECOMMENDATION

Strong recommendation against intervention	the Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	Strong recommendation for the intervention
0	0	0		0

#### **CONCLUSIONS**

#### Recommendation

In MDR/RR-TB patients on longer regimens, a total treatment duration of 18-20 months is suggested for most patients, and the duration may be modified according to the patient's response to therapy (conditional recommendation, very low certainty in the estimates of effect).

# **Justification**

Treatment of MDR-TB in adults and children with longer second-line regimens has been known to increase the likelihood of cure and lower the risk of chronicity and death (2),(3),(4),(5). The duration of effective treatment is an important consideration to maximise the likelihood of cure and to minimise the likelihood of poor treatment outcomes (such as treatment failure, relapse and death), serious adverse events and the emergence of additional resistance.

The evidence that informed this recommendation was a subset of the data from the 2018 IPD-MA (see Background section on first page). This relied heavily on observational studies with few studies conducted under randomized controlled conditions. Information on relapse was only available for 388 patients (with 9 relapse events) and is therefore very likely to be a highly selective patient group. One of the main concerns is that, particularly in patients on individualized regimens, treatment may have been prolonged if a patient was not improving clinically, and so patients who received longer therapy may be those who were sicker. Attempts to control for bias and confounding in the evidence review are unlikely to have adjusted for all factors. As a result, the quality of the evidence was judged to be very low.

Based on the results of the primary analyses, there was a marginally increased risk of treatment failure or relapse when the duration of MDR-TB treatment was 20-22 months (compared to 17.5-20 months), although the estimated risk was not statistically significant. There were no data available on harms associated with different durations of treatment. The GDG noted that the optimal treatment course is one that is effective in reducing treatment failure, relapse, death, harms and the emergence of additional drug resistance, while maintaining the likelihood of cure.

# **Subgroup considerations**

The recommendation also applies to patients previously treated with second line regimens and to XDR-TB patients.

MDR/RR-TB alone or with additional resistance: The results of the IPD-MA did not show any differences in treatment failure or relapse when comparing patients with MDR-TB with or without additional second-line drug resistance, including XDR-TB. An analysis of XDR-TB patients only, as part of the sensitivity analyses, showed no differences to the primary analyses. Despite this, the duration of treatment may need to be longer than 20 months overall in XDR-TB patients, subject to the clinical response to the treatment.

Children: Shortening total treatment duration to less than 18 months may be considered in the case of children without severe disease, generally determined on the basis of one or more of the following: the presence of cavities or bilateral disease on chest radiography or extrapulmonary forms of disease other than lymphadenopathy (peripheral nodes or isolated mediastinal mass without compression). The occurrence of advanced malnutrition (defined by syndrome or by metrics), of advanced immunosuppression, or positive TB bacteriology (smear, Xpert MTB/RIF, culture) may also be considered when determining disease severity.

Extrapulmonary TB and culture-negative laboratory confirmed TB: Extra-pulmonary MDR/RR-TB is generally treatable with the same combination of medicines and duration as pulmonary MDR/RR-TB (see also PICO question 2 regarding specific medicines for cerebral disease). Other durations of treatment may be appropriate for persons with culture negative laboratory confirmed TB. In such cases a total duration of treatment of 18-20 months is advised and the response should be monitored by clinical parameters other than specimen bacteriology. A negative culture result may reflect poor laboratory performance rather than true negativity of sputum, which underscores the importance of quality assurance in the laboratory.

## **Implementation considerations**

National TB programmes may choose to adhere to a fixed duration of regimen (e.g. 20 months) to facilitate the implementation of the recommendation.

This recommendation should be implemented in light of two other recommendations in the 2018 guidelines. The GDG conditionally recommends 15 to 17 months of treatment after culture conversion for most patients (see PICO question 6). The duration of treatment post culture conversion may be modified according to the patient's response to therapy and other risk factors for failure or relapse. The presence of XDR-TB and extent of disease were also factored into the analysis on the duration of treatment post culture conversion. Likewise, in longer regimens that include amikacin or streptomycin the recommendation on the duration of the intensive phase – usually 6-7 months - also applies (see PICO question 4)

## Monitoring and evaluation

Patients on longer MDR-TB treatment regimens need to be monitored for response to treatment, treatment failure and for safety using reasonable schedules of relevant clinical and laboratory testing (6),(7). Monitoring response to treatment is done through regular history taking, physical examination, chest radiograph and laboratory monitoring. Conversion of bacteriological status using smear microscopy or culture is an important means of assessing response and most patients are usually expected to have converted to negative within the first few months of starting treatment. Frameworks for the surveillance of bacteriological status, drug-resistance and the assignment of outcomes have been fairly standardized in past years (8). In contrast, the systematic monitoring of adverse events during and after the end of treatment needs to be strengthened in most TB programmes given the relative novelty of active pharmacovigilance within national TB programmes. The rationale for aDSM is largely supported by the increasing use worldwide of combinations of new and re-purposed medications in MDR-TB treatment regimens. The toxicity of certain agents (such as linezolid) is increased with duration of use and may limit their continued employment in a patient, and at times the complete cessation of treatment. The prospective collection of accurate data for key variables at case-based level using an electronic register is strongly advised in the best interests of the individual patient and to inform local and global policy revisions (9).

## Research priorities

In the process of summarizing the available evidence, the reviews undertaken for this update revealed a number of gaps in current knowledge about critical areas for the treatment for MDR/RR-TB.

Amongst these, the GDG remarked that further research is required to determine the optimal and minimum duration of treatment for MDR/RR-TB for culture negative patients, children and patients with extrapulmonary forms of MDR/RR-TB. The GDG noted that data from randomised controlled trials are expected in the next few years which could provide evidence in support of the optimization of total duration of treatment for different MDR/RR-TB subgroups.

### REFERENCES SUMMARY

- 1. Guidelines for the programmatic management of drug-resistant tuberculosis, 2011 Update. (WHO/HTM/TB/2011.6) [Internet]. Geneva, World Health Organization. 2011. Available from: http://whqlibdoc.who.int/publications/2011/9789241501583\_eng.pdf
- 2. Ahuja SD, Ashkin D, Avendano M, Banerjee R, Bauer M, Bayona JN, et al. Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: an individual patient data meta-analysis of 9,153 patients. PLoS Med. 2012;9(8):e1001300.
- 3. Ahmad N, Ahuja SD, Akkerman OW, Alffenaar J-WC, Anderson LF, Baghaei P, et al. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. The Lancet. 2018 Sep;392(10150):821–34.
- 4. Harausz EP, Garcia-Prats AJ, Law S, Schaaf HS, Kredo T, Seddon JA, et al. Treatment and outcomes in children with multidrug-resistant tuberculosis: A systematic review and individual patient data meta-analysis. Metcalfe JZ, editor. PLOS Med. 2018 Jul 11;15(7):e1002591.
- 5. Seddon JA, Hesseling AC, Godfrey-Faussett P, Schaaf HS. High treatment success in children treated for multidrug-resistant tuberculosis: an observational cohort study. Thorax. 2014 May 1;69(5):458-64.
- 6. Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis. (WHO/HTM/TB/2014.11) [Internet]. Geneva, World Health Organization; 2015. Available from: http://apps.who.int/iris/bitstream/10665/130918/1/9789241548809\_eng.pdf
- 7. Active tuberculosis drug-safety monitoring and management (aDSM). Framework for implementation. (WHO/HTM/TB/2015.28) [Internet]. Geneva, World Health Organization; 2015. Available from: http://apps.who.int/iris/bitstream/10665/204465/1/WHO HTM TB 2015.28 eng.pdf
- 8. Definitions and reporting framework for tuberculosis 2013 revision (WHO/HTM/TB/2013.2) [Internet]. Geneva, World Health Organization; 2013. Available from: http://apps.who.int/iris/bitstream/10665/79199/1/9789241505345 eng.pdf
- 9. Electronic recording and reporting for tuberculosis care and control (WHO/HTM/TB/2011.22) [Internet]. Geneva, World Health Organization; 2012. Available from: http://whqlibdoc.who.int/publications/2012/9789241564465 eng.pdf

## **PICO QUESTION 6**

In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines what is the minimum duration of treatment after culture conversion that is more likely to improve outcomes?

POPULATION: Patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens conforming to WHO guidelines

**INTERVENTION:** A treatment duration of 15-17 months after culture conversion

**COMPARISON:** A shorter (12-15 months) or longer (17-19 months) duration after culture conversion

**MAIN OUTCOMES:** Treatment failure or relapse versus treatment success

**SETTING:** Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in low-, middle- and high-income settings, using hospital or ambulatory models of care.

> Clinicians often demand guidance on the duration of MDR-TB treatment required once the patient is no longer culture positive. The findings of the analyses for PICO question 6 are expected to influence the continued validity in their present form of the expert opinion on the duration of MDR-TB treatment after culture conversion provided in the Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis(1). The duration of the intensive phase of longer regimens that include an injectable

agent is addressed in PICO question 4 while the total duration of treatment is addressed in PICO question 5.

In the Guidelines for the Management of Drug-Resistant Tuberculosis: Emergency Update 2008, it was proposed that the duration of treatment for MDR-TB be based on the use of a parenteral agent for a minimum of 6 months and at least 4 months past culture conversion, and a minimum total length of treatment of 18 months after culture conversion (2). The evidence reviews undertaken for the Guidelines for the Programmatic Management of Drug-Resistant Tuberculosis: 2011 Update could not inform whether a minimum length of intensive phase or total treatment after conversion was a determinant of treatment outcome (3),(4),(5),(6). Given significant changes introduced to the longer regimens used worldwide, with the global expansion of the use of later-generation fluoroquinolones, bedaquiline and longer intensive phase, and given that the 2018 IPD contains datasets from a wide geographical cross-section of countries, it was deemed timely to review evidence for this question to determine the optimal duration of treatment post culture conversion.

This PICO question was informed by an analysis of 4,175 patients from 39 observational studies, a subset of the 13,104 patients from 53 studies and 40 countries compiled for the 2018 guidelines update. All but 3 of the 4,175 patients were on individualized regimens. The reasons for exclusion of 8,929 records from the main dataset were as follows: outcome reported as lost to follow up: 2,261; outcome reported as died: 2,043; treatment duration not reported: 230; culture information not reported: 1,945; baseline culture negative: 754, patient never culture converted: 426; number of effective drugs less than five or less than four plus pyrazinamide: 1215; treatment duration less than six months: 4; treatment duration greater than or egual to 36 months: 49; culture converted post treatment: n=2. There was limited information on relapse; 81% of patients with information on treatment failure/ relapse were excluded due to the highly selective nature of the analyses. (GRADE summary of evidence table in Annex 8)

**PERSPECTIVE:** 

**BACKGROUND:** 

## **ASSESSMENT**

## Problem

Is the problem a priority?		
JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
O No O Probably no O Probably yes ● Yes O Varies O Don't know	About 558,000 new cases of rifampicin- and multidrug-resistant TB (MDR/RR-TB) cases are estimated to emerge worldwide each year and 230,000 cases die. Globally only about one fourth of newly emergent MDR-TB patients have been reported to start a second-line TB treatment annually in recent years.  Outcomes of MDR-TB treatment on a global level are poor with much loss to follow up and death, and 55% of cases reported to have a successful outcome at the end of treatment. The need for better treatment for these patients is thus a priority in an effort to save lives and reduce transmission and chronicity. The provision of effective treatment to optimise the likelihood of relapse free cure is also a priority, and the duration of treatment post culture conversion is an important consideration.	MDR-TB is a global challenge and access to treatment often problematic, with regimens typically being long, toxic and expensive.  There is clearly an interest in reducing the duration of treatment, simplifying the administration of the regimen, and providing patients with a safer combination of medicines that can cure the large majority of cases.

## **Desirable Effects**

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE					ADDITIONAL CONSIDERATIONS	
o Small o Moderate o Large o Varies o Don't know	participants (studies)	participants of (studies)	of the effect evidence (95%	effect (95%	Anticipated absolute effects* (95% CI)		The reference category considered by the GDG was 15-17 months post culture conversion. It was compared to 12-15 months post culture conversion and 17-19 months post culture conversion. The desirable effects were no different when comparing 15-17 months to 17-19 months post culture conversion.
		(GRADE) CI)	CI)	Risk with shorter or longer duration	Risk difference with a treatment duration of 15-17 months after culture conversion		
			ФООО VEDV	<b>aOR 5.5</b> (0.9 to	Study population		
	VLKI 22.7	33.7)	8 per 100	6 more per 100 (2 more to 11 more)			

17 Months Post-Culture Conversion					
Treatment 165 Failure/Relapse (39	39	ФФОО LOW	(0.4 to	Study pop	ulation
	oservational cudies)		3.7)	2 per 100	O fewer per 100 (1 fewer to 5 more)

## **Undesirable Effects**

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE						ADDITIONAL CONSIDERATIONS
o Large o Moderate o Small o Trivial o Varies • Don't know	Outcomes	participants		effect (95%	Anticipated absolute effects* (95% CI)		Apart from the information on the undesirable treatment outcomes of failure/ relapse, there was no information about harms available to the GDG. Therefore the GDG judged that the undesirable effects were not known, resulting in a judgment of Don't know.
		Follow up		Risk with shorter or longer duration	Risk difference with a treatment duration of 15-17 months after culture conversion		
	Treatment Failure/Relapse	Treatment Failure/Relapse vs. Treatment Success: Treatment Duration of 12- 15 Months Post-Culture Conversion vs. Treatment Duration of 15- 17 Months Post-Culture Conversion	⊕○○○ VERY	(0.9 to 33.7)	Study population		
	Success: Treatment Duration of 12- 15 Months Post-Culture Conversion vs. Treatment Duration of 15- 17 Months Post-Culture				8 per 100	6 more per 100 (2 more to 11 more)	
	Treatment Failure/Relapse	I	$  \bigcup_{i \in W} \bigcup_{j \in V}   (0.$	<b>aOR 1.2</b> (0.4 to	Study population		
	vs. Treatment Success: Treatment Duration of 17- 19 Months Post-Culture Conversion vs. Treatment Duration of 15- 17 Months	observational studies)		3.7)	2 per 100	O fewer per 100 (1 fewer to 5 more)	

	Post-Culture Conversion  a. Fixed effects aOR reported due to non-convergence of random-effects model.	
Certainty of evidence What is the overall certainty of the evidence of	effects?	
JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
Very low     Cow     Moderate     High     No included studies	These data were derived from an IPD-MA which comprised a total of 4175 patients for the primary analyses, from 39 observational studies. These 4175 patients were a sub-set of the individual patient dataset used for PICO 2, which comprised 13,104 patients overall, from 53 studies and 40 countries. The certainty of the evidence was judged to be very low.	The certainty of the evidence is very low based on the fact that the majority of studies included in the IPD are observational. Further, for the comparison of the outcomes of 12-15 months post culture conversion vs. 15-17 months post culture conversion, a fixed effects meta-analysis was used for the adjusted odds ratio (as the random effects model did not converge) resulting in the certainty of evidence being further downgraded for imprecision. The for the analyses of the 2018 IPD data were matched (exactly) for resistance to a fluoroquinolone, resistance to a second line injectable agent and income level (according to the World Bank Atlas method). In addition, propensity score matching was used for a number of other important covariates. Therefore, efforts were made to adjust for potential confounders, however residual confounding is possible.
Values Is there important uncertainty about or variability	ty in how much people value the main outcomes?	
JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
Important uncertainty or variability     Possibly important uncertainty or variability     Probably no important uncertainty or	No research evidence was identified.	Although no research evidence was identified the GDG concluded that there would be no important uncertainty or variability in how much people would value the outcomes.

variability

• No important uncertainty or variability

## **Balance of effects**

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
O Favors the comparison O Probably favors the comparison O Does not favor either the intervention or the comparison Probably favors the intervention O Favors the intervention O Varies O Don't know		The desirable effects were regarded as trivial by the GDG as there was little change in treatment success when comparing 15-17 months to 17-19 months post culture conversion. The undesirable effects were not clearly directional. Therefore, the GDG judged that the balance of effects probably favoured the intervention (i.e. 15-17 months post culture conversion).

Resources required
How large are the resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Large costs o Moderate costs o Negligible costs and savings o Moderate savings o Large savings ● Varies o Don't know	No research evidence was identified.	Although no research evidence was identified, the GDG considered two durations post culture conversion for this PICO with 15-17 months post culture conversion considered as the reference group. Given that one duration was shorter (i.e. 12-15 months) and that one was longer (i.e. 17-19 months) the GDG considered that the costs may vary and that they are dependent on the duration of treatment. Therefore the judgment was considered as: Varies.

# Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
<ul> <li>Very low</li> <li>Low</li> <li>Moderate</li> <li>High</li> <li>No included studies</li> </ul>	No research evidence was identified.	Although no research evidence was identified, the GDG acknowledged that the certainty of the evidence of required resources was very low.

## **Cost effectiveness**

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
O Favors the comparison O Probably favors the comparison O Does not favor either the intervention or the comparison O Probably favors the intervention O Favors the intervention O Varies  No included studies	No research evidence was identified.	

**Equity**What would be the impact on health equity?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Reduced o Probably reduced o Probably no impact ● Probably increased o Increased o Varies o Don't know	No research evidence was identified.	Although no research evidence was identified it was assumed that patients will receive the correct duration of treatment if the duration of treatment after culture conversion is known (i.e. patients will be more likely to receive effective treatment and will forego sub-optimal treatment, leading to the GDG to make a judgment of probably increased, in relation to equity).

## Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o No o Probably no ● Probably yes		Although no research evidence was identified, the GDG thought that policy makers may find shorter and effective treatment regimens more acceptable.
o Yes o Varies o Don't know		

## **Feasibility**

Is the intervention feasible to implement?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o No o Probably no o Probably yes ● Yes		Although no research evidence was identified the GDG thought that a shorter and effective regimen is feasible to implement given that this duration of regimen is already being implemented

o Varies		
o Don't know		

in the majority of countries where MDR/RR-TB is being treated globally.

## **SUMMARY OF JUDGEMENTS**

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	Yes		Varies	Don't know
DESIRABLE EFFECTS	Trivial	Small	Moderate	Large		Varies	Don't know
UNDESIRABLE EFFECTS	Large	Moderate	Small	Trivial		Varies	Don't know
CERTAINTY OF EVIDENCE	Very low	Low	Moderate	High			No included studies
VALUES	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability			
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	Don't know
RESOURCES REQUIRED	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	Low	Moderate	High			No included studies
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	No included studies
EQUITY	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know
ACCEPTABILITY	No	Probably no	Probably yes	Yes		Varies	Don't know
FEASIBILITY	No	Probably no	Probably yes	Yes		Varies	Don't know

### TYPE OF RECOMMENDATION

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	Strong recommendation for the intervention
0	0	0	•	0

## **CONCLUSIONS**

### Recommendation

In MDR/RR-TB patients on longer regimens, a treatment duration of 15 to 17 months after culture conversion is suggested for most patients, and the duration may be modified according to the patient's response to therapy (conditional recommendation, very low certainty in the estimates of effect).

## **Justification**

Treatment of MDR-TB in adults and children with longer second-line regimens has been known to increase the likelihood of cure and lower the risk of chronicity and death (2),(3),(4),(5). The duration of effective treatment post culture conversion is an important consideration to maximise the likelihood of cure and to minimise the likelihood of poor treatment outcomes (such as treatment failure, relapse and death), serious adverse events and the emergence of additional resistance.

The evidence that informed this recommendation was a subset of the data from the 2018 IPD-MA (see Background section on first page). This relied heavily on observational studies with few studies conducted under randomized controlled conditions. One of the main concerns is that treatment may have been prolonged if a patient was not improving clinically, and so patients who received longer therapy may be those who were sicker. Attempts to control for bias and confounding in the evidence review are unlikely to have adjusted for all factors. Another limitation of these data was that the availability of information on relapse was sparse due to the highly selective nature of the analyses. As a result, the quality of the evidence was judged to be very low.

Based on the results of the primary analyses, there was an increased risk of treatment failure or relapse when the duration of treatment post culture conversion was 12-15 months (compared to 15-17 months), which was statistically significant, but no difference in risk if the regimen was extended to 17-19 months post culture conversion. There were no data available on harms associated with different durations of treatment. The GDG noted that the optimal treatment course is one that is effective in reducing treatment failure, relapse, death, harms and the emergence of additional drug resistance, while maintaining the likelihood of cure.

## **Subgroup considerations**

The recommendation also applies to patients previously treated with second line regimens, XDR-TB patients and to patients with extensive disease.

MDR/RR-TB alone or with additional resistance: The results of the IPD meta-analysis did not show any differences in treatment failure or relapse when comparing patients with MDR-TB with or without additional resistance (e.g. MDR-TB vs XDR-TB). The duration of treatment may need to be longer than 20 months overall in total for patients with additional resistance, although this may also depend on the clinical response to treatment.

Children: This recommendation may not apply to children who commonly do not have a bacteriological diagnosis. In this case the monitoring of response is determined by other indicators.

Extra-pulmonary TB: This recommendation does not apply to persons with extra pulmonary TB, in whom response and treatment duration are determined by other criteria for treatment response.

Culture-negative laboratory confirmed TB: This recommendation does not apply and total duration should follow the recommendation for PICO question 5. It should be noted that a negative culture may not necessarily reflect the true status of the patient but rather poor transportation or storage of a sample or else contamination of the specimen. This underscores the importance of the quality assurance of laboratory work.

Persons with extensive TB disease: The duration of treatment post culture conversion may be modified according to the patient's response to therapy (e.g. culture conversion before 2 months of treatment) and other risk factors for treatment failure or relapse. This should be considered in patients with extensive TB disease.

## **Implementation considerations**

The duration may be shorter than the recommended total duration if culture conversion occurs before 2 months of treatment and in the absence of risk for failure or relapse. National TB programmes may choose a fixed duration (e.g. 16 months) for implementation purposes. Some countries experience difficulties with the implementation and quality assurance of sputum culture, which impacts upon this recommendation as it is dependent on access to culture. The yield of smear microscopy and culture also depend on the quality of the sputum produced, so care should be taken to obtain adequate specimens and transporting them to the laboratory according to standard procedures to maintain viability of the bacilli to get a valid culture result.

## **Monitoring and evaluation**

Patients on longer MDR-TB treatment regimens need to be monitored for response to treatment, treatment failure and for safety using reasonable schedules of relevant clinical and laboratory testing (1),(7). Monitoring response to treatment is done through regular history taking, physical examination, chest radiograph and laboratory monitoring. Conversion of bacteriological status using smear microscopy or culture is an important means of assessing response and most patients are usually expected to have converted to negative within the first few months of starting treatment. Sputum culture is much more sensitive than sputum smear to detect ongoing active disease and/or treatment failure. Frameworks for the surveillance of bacteriological status, drug-resistance and the assignment of outcomes have been fairly standardized in past years (8). In contrast, the systematic monitoring of adverse events during and after the end of treatment needs to be strengthened in most TB programmes given the relative novelty of active pharmacovigilance within national TB programmes. The rationale for aDSM is largely supported by the increasing use worldwide of combinations of new and re-purposed medications in MDR-TB treatment regimens. The toxicity of certain agents (such as linezolid) is increased with duration of use and may limit their continued employment in a patient, and at times the complete cessation of treatment. The prospective collection of accurate data for key variables at case-based level using an electronic register is strongly advised in the best interests of the individual patient and to inform local and global policy revisions (9).

## **Research priorities**

In the process of summarizing the available evidence, the reviews undertaken for this update revealed a number of gaps in current knowledge about critical areas for the treatment for MDR/RR-TB.

Amongst these, the GDG remarked that clinical trials that include MDR-TB regimens with different durations of treatment are a priority. The GDG noted that different phenotypes may need different types of therapy, different durations of therapy overall and different durations of therapy beyond culture conversion, which should be the subject of future research.

### REFERENCES SUMMARY

- 1. Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis. (WHO/HTM/TB/2014.11) [Internet]. Geneva, World Health Organization; 2015. Available from: http://apps.who.int/iris/bitstream/10665/130918/1/9789241548809 eng.pdf
- 2. Guidelines for the programmatic management of drug-resistant tuberculosis, Emergency update 2008. (WHO/HTM/TB/2008.402) [Internet]. Geneva, World Health Organization. 2008. Available from: http://whqlibdoc.who.int/publications/2008/9789241547581 eng.pdf
- 3. Guidelines for the programmatic management of drug-resistant tuberculosis, 2011 Update. (WHO/HTM/TB/2011.6) [Internet]. Geneva, World Health Organization. 2011. Available from: http://whqlibdoc.who.int/publications/2011/9789241501583 eng.pdf
- 4. Ahuja SD, Ashkin D, Avendano M, Banerjee R, Bauer M, Bayona JN, et al. Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: an individual patient data meta-analysis of 9,153 patients. PLoS Med. 2012;9(8):e1001300.
- 5. Orenstein EW, Basu S, Shah NS, Andrews JR, Friedland GH, Moll AP, et al. Treatment outcomes among patients with multidrug-resistant tuberculosis: systematic review and meta-analysis. Lancet Infect Dis. 2009 Mar;9(3):153–61.
- 6. Johnston JC, Shahidi NC, Sadatsafavi M, Fitzgerald JM. Treatment Outcomes of Multidrug-Resistant Tuberculosis: A Systematic Review and Meta-Analysis. PLoS ONE. 2009 Sep 9;4(9):e6914.
- 7. Active tuberculosis drug-safety monitoring and management (aDSM). Framework for implementation.(WHO/HTM/TB/2015.28) [Internet]. Geneva, World Health Organization; 2015. Available from: http://apps.who.int/iris/bitstream/10665/204465/1/WHO HTM TB 2015.28 eng.pdf
- 8. Definitions and reporting framework for tuberculosis 2013 revision (WHO/HTM/TB/2013.2) [Internet]. Geneva, World Health Organization; 2013. Available from: http://apps.who.int/iris/bitstream/10665/79199/1/9789241505345 eng.pdf
- 9. Electronic recording and reporting for tuberculosis care and control (WHO/HTM/TB/2011.22) [Internet]. Geneva, World Health Organization; 2012. Available from: http://whqlibdoc.who.int/publications/2012/9789241564465\_eng.pdf

## **PICO QUESTION 7**

**BACKGROUND:** 

In patients with MDR/RR-TB treated with longer or shorter regimens composed in accordance with WHO guidelines, is monitoring using monthly cultures, in addition to smear microscopy, more likely to detect non response to treatment?

POPULATION: Patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens conforming to WHO guidelines

INTERVENTION: Monthly culture and sputum smear microscopy

COMPARISON: Monthly sputum smear microscopy alone

PURPOSE OF THE TEST:

Achieving sustained bacteriological conversion from positive to negative is widely used to assess response to treatment in both drug-susceptible and drug-resistant tuberculosis. Culture is a more sensitive test for conversion to negative, or reversion to positive, than direct microscopy of sputum and other biological specimens. Culture also facilitates phenotypic testing for drug-

susceptibility testing, a critical consideration in tuberculosis diagnostics.

ROLE OF THE TEST: Culture and microscopy are widely used in tuberculosis treatment programmes both to diagnose new cases before start of treatment and to monitor regimen performance.

LINKED TREATMENTS: Longer or shorter MDR-TB regimens composed in accordance with WHO guidelines

ANTICIPATED OUTCOMES: Culture conversion by 6 months; Treatment failure or relapse; Acquisition (amplification) of drug resistance

SETTING:

Treatment of MDR/RR-TB patients with or without additional resistance using longer regimens in hospital or ambulatory models of care in low-, middle- and high-income countries.

PERSPECTIVE: The findings of these analyses are expected to influence the continued validity in its present form of a conditional recommendation made by WHO in 2011 on the use of sputum culture and

smear microscopy for the monitoring of patients with MDR-TB, during treatment. The data analysis was done to (i) compare the performance of the two methods in terms of

sensitivity/specificity and (ii) to assess the minimum frequency of testing needed in order not to delay unnecessarily any revision of the treatment.

Culture is a more sensitive test for bacteriological confirmation of tuberculosis than direct microscopy of sputum and other biological specimens. Culture also facilitates phenotypic testing for drug-susceptibility testing, a critical consideration in tuberculosis diagnostics. However, performing culture requires a better equipped laboratory to limit cross contamination, ensure proper bacterial growth and match other quality standards. Apart from the resource requirements culture results entail a significant delay of weeks or months, contrasting markedly with the relative immediacy of direct microscopy. While molecular techniques can now provide a rapid and reliable diagnosis they cannot replace culture or microscopy for the monitoring of bacteriological status during the course of treatment.

In 2011, WHO guidelines recommended that patients on MDR-TB treatment be monitored using both sputum smear microscopy and culture rather than sputum smear microscopy alone. (1) Performing both monthly sputum smear microscopy and culture was the best strategy in identifying failure earlier. The evidence underpinning the 2011 recommendations was derived from 10 published observational studies for which individual MDR-TB patient data were available to analyse the risk of failure using monthly culture (as reference) to alternative monitoring strategies. The estimates of effect were judged to be of very low certainty by the GDG advising WHO. In 2018, WHO on the advice of a GDG convened to revise its MDR/RR-TB treatment guidelines, reopened the discussion on the choice and frequency of testing to monitor bacteriological response to MDR-TB treatment. This question was prioritised given significant changes to MDR-TB management in the last decade, including the more widespread use of later generation fluoroquinolones, bedaquiline, linezolid and other effective agents with a longer intensive phase being recommended (about 8 months), implying that reversion to positive during the continuation phase would be less common; increased use of the 9-month shorter MDR-TB regimen: and increased use of liquid culture.

To address PICO question 7 the data were derived from the South African dataset, which comprised 26,522 patients overall. Of these, 22,760 records were excluded from the dataset, for the following reasons: 11,236 had a treatment outcome of death or loss to follow up; 698 had a treatment outcome of success but had less than 17.5 months of treatment; 1,357 had fewer than six culture samples recorded; 1,632 had no baseline culture recorded; 2,502 were baseline culture negative; 2,920 were smear negative at baseline or had a missing smear at baseline and 2,415 had insufficient smear data to match the culture data. This left 3,762 patients (1.8% <15 years) treated with longer MDR-TB regimens between 2010 and 2015 who had both monthly smear and culture data throughout treatment to address PICO question 7. The analysis focused only whether monthly culture vs. monthly smear microscopy or culture every two months is needed not to miss treatment failure in MDR/RR-TB patients on treatment. The odds of treatment failure in patients who do not convert at 6 months or later was also discussed. The data could not address the outcome on acquisition (amplification) of drug resistance or on relationship with failure in the 9-12 month shorter MDR-TB regimen envisaged in the original PICO 7. (GRADE summary of evidence table in Annex 8)

subgroups:

none (nearly all patients were adults with pulmonary disease on longer MDR-TB regimens composed in accordance with WHO guidelines; HIV prevalence was 60% and therefore much higher than in many settings outside of this region)

## **ASSESSMENT**

## Problem

Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o No o Probably no o Probably yes	About 558,000 new cases of rifampicin- and multidrug-resistant TB (MDR/RR-TB) cases are estimated to emerge worldwide each year and 230,000 cases die (WHO, 2018). Globally only about one fourth of newly emergent MDR-TB patients have been reported to start a second-line TB treatment annually in recent years.	MDR-TB is a global challenge and access to treatment often problematic, with regimens typically being long, toxic and expensive.
Yes     Varies     Don't know	Outcomes of MDR-TB treatment on a global level are poor with much loss to follow up and death, and only about 55% of cases reported to have a successful outcome at the end of treatment. The need for better treatment for these patients is thus a priority in an effort to save lives and reduce transmission and chronicity.	There is clearly an interest in reducing the duration of treatment, simplifying the administration of the treatment regimen, and providing patients with a safer combination of medicines that can cure the large majority of cases.

## **Test accuracy**

How accurate is the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
<ul> <li>Very inaccurate</li> <li>Inaccurate</li> <li>Accurate</li> <li>Very accurate</li> <li>Varies</li> </ul>	TB culture is considered the gold standard test for diagnosing TB. Culture also facilitates phenotypic testing for drug-susceptibility testing, a critical consideration in TB diagnostics. Culture is a more sensitive test for conversion from positive to negative, or reversion to positive, than direct smear microscopy of sputum (estimated sensitivity of 80%) and other biological specimens.	The accuracy of sputum culture (or smear microscopy) may be affected by a number of factors including the quality of the sputum specimen, quality assurance issues in the laboratory and other issues related to the handling and storage of the specimen.
o Don't know		

## **Desirable Effects**

How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE		ADDITIONAL CONSIDERATIONS			
o Trivial o Small ■ Moderate o Large o Varies o Don't know			results per s tested (95% I)	Nº of Certainty of		Concomitant use of sputum smear microscopy and culture test results helps identify patients whose bacteriology remains positive or reverts to positive following initial conversion to negative. This helps clinicians to identify patients likely to fail their treatment and instituting
	Test result Prevalence 3%			participants (studies)	the evidence (GRADE)	infection control measures in a timely manner. Additional benefits would be expected from reduced transmission
		monthly culture	monthly smear			and development of resistance as well as appropriate changes to treatment regimens, but these were not explicitly addressed by the analysis. Given the higher sensitivity of monthly culture when compared to monthly
	True positives patients with	28 (26 to 29)	15 (13 to 18)	3762 (1)	⊕⊕⊕○ MODERATEª,b	smear microscopy (0.93 vs. 0.51), monthly culture increases the detection of patients with a true positive
	treatment failure	13 more TP in monthly culture				bacteriological result by 13 per 1,000 patients when compared with sputum smear microscopy alone. It also helps reduce on false negatives.

False negatives patients incorrectly	2 (1 to 4)	15 (12 to 17)		
classified as not having treatment failure	13 fewer FN in culture	monthly		
True negatives patients without treatment failure	940 (933 to 944)	957 (953 to 961)	3762 (1)	⊕⊕⊕⊖ MODERATE <sup>a,b</sup>
	17 fewer TN in monthly culture			
False positives patients incorrectly	30 (26 to 37)	13 (9 to 17)		
classified as having treatment failure	17 more FP in monthly culture			

Collecting the sample to carry out the test comes at no added effort on the patient.

- a. The population included had an HIV-positive prevalence of 60.2%.
  b. The GDG considered that precision was not seriously influenced even if analysis refers to data from one setting, South Africa.

Odds of failure (95% CL) in patients without sputum conversion by the end of successive months of treatment

		Month						
	1	2	3	4	5	6	7	8
According to Culture	3.6	4.1	5.2	7.4	10.3	16.4	24.7	44.5
Culture	(2.11, 5.97)	(2.76, 6.09)	(3.55, 7.55)	(5.00, 10.8)	(6.88, 15.38)	(10.72, 25)	(15.53, 39.20)	(26.53, 74.46)
According to	1.9	2.7	3.2	4.2	6.8	10.4	16.5	28.9
Smear	(1.27, 2.73)	(1.82, 3.88)	(2.11, 4.73)	(2.69, 6.48)	(4.19, 10.97)	(6.00, 17.92)	(9.15, 29.77)	(14.87, 56.14)

	lesira			
linc	ACITA	n a	L TTC	ctc
UIIU		DIE.		

How substantial are the undesirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Large		The higher sensitivity of culture with respect to sputum
<ul> <li>Moderate</li> </ul>		smear microscopy is that it leads to more false positives
o Small		(17 more per 1,000 patients). These patients may be
o Trivial		overtreated on an incorrect premise of non-response.
o Varies		Given the slightly lower specificity of monthly culture
o Don't know		when compared with sputum smear microscopy alone
		(0.97 vs. 0.99), monthly culture testing tends to identify
		more patients who do not actually sustain a treatment
		failure (17 fewer true negatives per 1,000 patients). The
		implication here is that treatment may be prolonged in the
		case of false positivity or missed true negativity; these
		tests may need to be repeated if there is suspicion of nor
		response, but the added inconvenience on the patient an
		programme is considered relatively small given that takin
		sputum and many other biological specimens is non-
		invasive and routine practice in many programmes. A
		bigger concern would be if more patients may be falsely
		assumed to be doing well when in fact they may have
		incipient failure. In a setting where testing is repeated at
		monthly intervals a single false positive or false negative
		test result is unlikely to prove harmful to the patient
		because treatment decisions usually rely upon at least to
		consecutive positive results (to denote prolonged
		positivity or reversion) and the effect of one spurious
		result would only last until the test repeated one month
		later is reported (see also further down).

Certainty	v of the evic	dence of to	est accuracy
oci tallit	y or the con-		

What is the overall certainty of the evidence of test accuracy?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low o Low ■ Moderate o High o No included studies		Although no research was identified, culture is considered the gold standard test for bacteriological confirmation of TB and had a better sensitivity over sputum smear microscopy to detect treatment failure when used for treatment monitoring. There was overall certainty about the risk of missing or delaying the detection of failure if smear microscopy alone was used instead of culture. If the assignment of treatment failure in this dataset relied more on smear microscopy than culture then bias against culture could have been introduced (which could explain the lower specificity of culture that was observed).

## **Certainty of the evidence of test's effects**

What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low o Low • Moderate o High o No included studies		One of the direct benefits of the test is its sensitivity to diagnose TB. In contrast to smear microscopy culture confirms the viability of the organisms detected and helps distinguish from non tuberculous mycobacteria. Culture also facilitates phenotypic drug-susceptibility testing - a critical component of care for persons with MDR/RR-TB. The test itself has negligible adverse effects when carried out according to standard operating procedures being a routine type of analysis done in TB patients and usually not requiring any invasive methods to obtain sputum, the most common specimen used for bacteriological examination.

# Certainty of the evidence of management's effects What is the overall certainty of the evidence of effects of the management that is guided by the test results?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low o Low o Moderate o High ● No included studies		When conducted in quality assured laboratories the risks for contamination or no growth because of improper sample handling are minimised. An isolated false positive or false negative test result is unlikely to prove harmful to the patient because treatment decisions usually rely upon at least two consecutive positive results (to denote prolonged positivity or reversion) and a spurious result would be corrected when the result of the test repeated one month later becomes available. This reinforces the importance of more frequent testing.

# Certainty of the evidence of test result/management How certain is the link between test results and management decisions?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low o Low o Moderate o High • No included studies		In TB treatment practice, the result of culture testing is one of the main considerations to determine treatment decisions such as the start of continuation phase (if an injectable agent is used), a change of regimen, termination of treatment and the assignment of cure or treatment failure as an outcome. Culture is also important because it confirms TB definitively at start of treatment or upon relapse and the positive sample can be subsequently referred for phenotypic or genotypic drug susceptibility testing, which help determine which regimen can be used or which drugs are likely to be effective.

## **Certainty of effects**

What is the overall certainty of the evidence of effects of the test?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
Low     Moderate     High     No included studies	The dataset used to address this question came from a subset of the programmatic, observational patient cohorts reported to WHO by South Africa for the guidelines revision. It was composed of data from 3,762 adult patients with pulmonary forms of MDR/RR-TB who had information on monthly sputum smear microscopy and culture throughout treatment with longer regimens in recent years. It thus differed from the main individual patient dataset merged across 53 studies in up to 40 countries and used to analyse for PICO 2. The overall certainty of the evidence was classified as moderate. The findings concur with those of two studies of MDR-TB patients treated with earlier regimens, which also concluded in favour of more frequent culture testing (2),(3).	The effects observed may vary in patients or populations with a profile markedly different from the one included in the analysis, such as low HIV prevalence settings, children, patients with extrapulmonary forms of disease or those treated with the shorter regimen. The 3,762 patients included in the analysis had very similar clinical characteristics to the 22,760 individuals excluded although they were slightly less likely to be HIV co-infected, have a history of previous treatment or to have second-line drug resistance. On the other hand, the rate of failure in those included in the analysis was only 3% compared to 12.7% of those excluded from the analysis. It should be expected that culture will always be a more sensitive test of bacterial positive status.

## **Values**

Is there important uncertainty about or variability in how much people value the main outcomes?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Important uncertainty or variability o Possibly important uncertainty or variability o Probably no important uncertainty or variability • No important uncertainty or variability	No research evidence was identified.	Although no research evidence was identified the GDG concluded that there would be no important uncertainty or variability in how much people would value the main outcomes of interest.  A high value was placed by the GDG on preventing treatment failure, relapse or death; maximising cure; preventing acquisition of additional resistance, all of which were classified as Critical or Important. Decreasing the transmission of MDR-TB that could result from its delayed diagnosis was also considered very relevant to this question. It is expected that most practitioners would prefer to base decisions on culture if this is possible and the GDG appreciates that the main limitation to perform culture more often is usually the availability of resources and, at times, the inability of the symptomless patient to produce a viable sputum sample.

_	_ 1	1		_ C	_ CC	ect	
-		2 r	CO	OT.	ΔTT	OCT	c
ъ,	<i>[ • ]</i>		119.5	UI.	СП	. – , – , –	-

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Favors the comparison o Probably favors the comparison		The main advantage of direct microscopy is that it can provide a result within hours in contrast to weeks or months for culture. This has important implications for immediate decisions on changes to treatment or separation of
o Does not favor either the intervention		patients to limit transmission. The advantage of culture is an improved accuracy of the test over smear microscopy
or the comparison		and a lower threshold to detect bacteria in the sputum sample. Practitioners in general would prefer to make
Probably favors the intervention     Favors the intervention		decisions on the basis of culture results. An isolated false positive or false negative culture test result is unlikely to prove harmful to the patient because treatment decisions usually rely upon at least two consecutive positive
o Varies		results (to denote prolonged positivity or reversion); a spurious result could thus be corrected when the result of
o Don't know		the test repeated one month later becomes available.

Resources required
How large are the resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Large costs o Moderate costs o Negligible costs and savings o Moderate savings o Large savings • Varies o Don't know	No specific studies done	Carrying out culture to high quality standards requires important investment in infrastructure, equipment, consumables and staff training. The test may only be carried out in a suitably-equipped laboratory, requiring more resources than those needed for smear microscopy. Once the setup is in place the unit costs of performing culture become much more affordable if the volume of testing is kept uniform and within the critical ranges of laboratory capacity, to avoid either overburdening the system or, conversely, periods of inactivity. High throughput systems using liquid culture are preferred. Solid culture methods are less expensive than liquid, but standardization of media is more difficult, more laborious and processing is much slower than liquid culture. Resources invested in improving access to culture would yield benefits in better detection of MDR-TB and case management.

# **Certainty of evidence of required resources**What is the certainty of the evidence of resource requirements (costs)?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Very low	No research evidence was identified.	
o Low		
o Moderate		
o High		
No included studies		

$\overline{}$			_ C	e -			ness
	OC.	т.	ΩТ	ТΟ	CTI	VA.	nacc
•	$\mathbf{u}$	ъ.	CI	_	чи	~~	

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Favors the comparison o Probably favors the comparison o Does not favor either the intervention or the comparison o Probably favors the intervention o Favors the intervention O Varies • No included studies	No research evidence was identified.	

**Equity**What would be the impact on health equity?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o Reduced o Probably reduced o Probably no impact o Probably increased o Increased • Varies o Don't know	No research evidence was identified.	Although no research was identified, it was acknowledged that solid and liquid culture methods are generally considered suitable for central reference laboratories (regional laboratories in large countries) or intermediate level laboratories, which may affect equity. Liquid culture methods are also more expensive that solid culture methods, but they increase the case yield by approximately 10% over solid media, and automated systems have the potential to reduce the diagnostic delay to days rather than weeks. (4) Access to liquid versus solid culture methods may also impact on health equity.  In a system with finite resources, a requirement to undertake culture more frequently than the current practice is likely to deviate more programmatic resources towards this function and may result in less funding available for other important laboratory activities (e.g. quality control, sample transportation or performance of surveys). This could impinge upon overall health equity if resources are redistributed to ensure better equipped, central diagnostic centres. When this recommendation is being considered the implementers need to plan carefully to ensure that sufficient funding is made available to develop the capacity as the workload increases.

## Acceptability

Is the intervention acceptable to key stakeholders?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
o No o Probably no ● Probably yes o Yes o Varies o Don't know		Culture is already being implemented in many countries globally and policy makers and clinicians are aware of its advantages and downsides. Producing more frequent sputum samples and delivering them to the diagnostic facility may impinge upon the acceptability of the recommendation by patients and health care staff. Patients who do not have or no longer produce sputum will not find the recommendation acceptable. However, in children unable to expectorate, gastric aspirates or induced sputa repeated at monthly frequency may not be acceptable. Nonetheless the recommendation reinforced a near-identical one made by WHO in 2011 on the basis of earlier analysis and modelling.

Feasibility Is the intervention feasible to implement?					
JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS			
o No o Probably no o Probably yes o Yes ● Varies o Don't know		The intervention can only be implemented if costs can be met, both in terms of increased real expense (e.g. staff, consumables, specimen transport) and increased need for facilities to carry out culturing. Culture is already being implemented in many countries globally. Solid and liquid culture methods are generally considered suitable for central reference laboratories (regional laboratories in large countries) or intermediate level laboratories. Solid culture methods are less expensive than liquid, but the results take longer to arrive (because of the slow growth of mycobacteria), which may limit their use in clincial decision-making. Culture may be unavailable in peripheral laboratories within countries, limiting its feasibility for many patients in a decentralised setting. Liquid culture increases the case yield by approximately 10% over solid media, and automated systems reduce the diagnostic delay to days rather than weeks. However, liquid systems are more prone to contamination, and the manipulation of large volumes of infectious material mandates appropriate additional biosafety measures.(4)			

## **SUMMARY OF JUDGEMENTS**

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	Yes		Varies	Don't know
TEST ACCURACY	Very inaccurate	Inaccurate	Accurate	Very accurate		Varies	Don't know
DESIRABLE EFFECTS	Trivial	Small	Moderate	Large		Varies	Don't know
UNDESIRABLE EFFECTS	Large	Moderate	Small	Trivial		Varies	Don't know
CERTAINTY OF THE EVIDENCE OF TEST ACCURACY	Very low	Low	Moderate	High			No included studies
CERTAINTY OF THE EVIDENCE OF TEST'S EFFECTS	Very low	Low	Moderate	High			No included studies
CERTAINTY OF THE EVIDENCE OF MANAGEMENT'S EFFECTS	Very low	Low	Moderate	High			No included studies
CERTAINTY OF THE EVIDENCE OF TEST RESULT/MANAGEMENT	Very low	Low	Moderate	High			No included studies
CERTAINTY OF EFFECTS	Very low	Low	Moderate	High			No included studies
VALUES	Important uncertainty or variability	Possibly important uncertainty or variability	Probably no important uncertainty or variability	No important uncertainty or variability			
BALANCE OF EFFECTS	Favors the comparison	Probably favors the	Does not favor either	Probably favors the	Favors the intervention	Varies	Don't know

	JUDGEMENT						
		comparison	the intervention or the comparison	intervention			
RESOURCES REQUIRED	Large costs	Moderate costs	Negligible costs and savings	Moderate savings	Large savings	Varies	Don't know
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	Low	Moderate	High			No included studies
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	Favors the intervention	Varies	No included studies
EQUITY	Reduced	Probably reduced	Probably no impact	Probably increased	Increased	Varies	Don't know
ACCEPTABILITY	No	Probably no	Probably yes	Yes		Varies	Don't know
FEASIBILITY	No	Probably no	Probably yes	Yes		Varies	Don't know

### TYPE OF RECOMMENDATION

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	Conditional recommendation for the intervention	Strong recommendation for the intervention
0	0	0	•	0

## **CONCLUSIONS**

### Recommendation

In MDR/RR-TB patients on longer regimens, the performance of sputum culture in addition to sputum smear microscopy is recommended to monitor treatment response (strong recommendation, moderate certainty in the estimates of test accuracy). It is desirable for sputum culture to be repeated at monthly intervals.

### **Justification**

Achieving sustained bacteriological conversion from positive to negative is widely used to assess response to treatment in both drug-susceptible and drug-resistant tuberculosis. Culture is a more sensitive test for conversion to negative, or reversion to positive, than direct microscopy of sputum and other biological specimens. Culture also facilitates phenotypic testing for drug-susceptibility testing, a critical consideration in tuberculosis diagnostics. In addition, culture conversion is used as a TB treatment indicator which allows clinicians to make decisions regarding treatment, such as the duration of treatment post culture conversion. Culture conversion may also be a prognostic marker of a successful treatment outcome.

To address PICO question 7 the data were derived from the South African dataset, which comprised 26,522 patients overall. Of these, 22,760 records were excluded from the dataset, for the following reasons: 11,236 had a treatment outcome of death or loss to follow up; 698 had a treatment outcome of success but had less than 17.5 months of treatment; 1,357 had fewer than six culture samples recorded; 1,632 had no baseline culture recorded; 2,502 were baseline culture negative; 2,920 were smear negative at baseline or had a missing smear at baseline and 2,415 had insufficient smear data to match the culture data. This left 3,762 patients (1.8% <15 years) treated with longer MDR-TB regimens between 2010 and 2015, who had both monthly smear and culture data throughout treatment to address PICO question 7. The analysis focused only whether monthly culture vs. monthly smear microscopy or culture every two months is needed not to miss treatment failure in MDR/RR-TB patients on treatment. The odds of treatment failure in patients who do not convert at 6 months or later was also discussed. The data could not address the outcome on acquisition (amplification) of drug resistance or on relationship with failure in the 9-12 month shorter MDR-TB regimen envisaged in the original PICO 7.

Based on results of these analyses the sensitivity of monthly culture to detect treatment failure was 93% compared to 51% for monthly sputum smear. Thirteen more per 1000 patients would have a true positive culture when comparing monthly culture to monthly smear. Based on the finding that the sensitivity of monthly culture is much higher than that of monthly smear to detect treatment failure, monthly culture was then compared to two monthly culture. The sensitivity of monthly culture to detect treatment failure was 93%, compared to 73% for two monthly culture. The overall quality of the evidence was classified as low although there was moderate certainty in the estimates of test accuracy.

## **Subgroup considerations**

The evidence informing this recommendation relates to an analysis performed on data of adult MDR-TB patients with pulmonary disease treated with contemporary longer regimens in recent years in South Africa. About 60% of these patients were HIV positive, a large proportion of whom was on treatment. The recommendation would apply to any longer regimen, regardless of the number of Group A, B or C agents used and whether an injectable (intensive phase) applied or not. The GDG considered that the findings may apply to other key patient subgroups.

Patients <15 years of age younger children with MDR/RR-TB commonly cannot produce sufficient sputum spontaneously to allow a bacteriological diagnosis (many are typically sputum smear microscopy negative). In these patients culture may be a more sensitive means to detect viable TB bacilli even if very few organisms are present in sputum or other samples, below the detection threshold of direct microscopy. However, in children unable to expectorate, gastric aspirates or induced sputa repeated at monthly frequency may not be acceptable.

Extrapulmonary disease is commonly paucibacillary and biological specimens may therefore contain few or no bacilli. In such a situation culture stands a better chance of detecting persistent disease although collection of samples often poses problems. Direct microscopy should still be attempted because it may determine positivity much faster than culture

HIV-negative individuals with TB typically have higher bacterial counts in sputum. In such a situation one may expect that the difference in test sensitivity between smear and culture would be less extreme given that

fewer patients would have sub-threshold bacterial counts. However, past studies on datasets from multiple sites in which HIV positivity was low reported findings that led to the WHO recommendation even in 2011 for joint use of both microscopy and culture, preferably every month.

Patients on the shorter MDR-TB regimen have a much shorter duration of intensive and total treatment, receive 7 drugs in the initial phase and, given a number of inclusion/exclusion criteria, usually have a more favourable prognostic score than the average MDR-TB treatment. The current analysis did not include patients treated with these regimens. It may thus be argued that in patients on the shorter MDR-TB regimen there are less grounds to support a recommendation for the use of monthly culture to monitor treatment. The GDG however proposes that programmes that implement this regimen aim for more frequent culture testing especially after the intensive phase in order to confirm bacteriological cure in patients who complete treatment without signs of failure. Any sign of recurrence after termination of treatment should also be investigated using both sputum microscopy and culture.

## Implementation considerations

Good quality sputum specimens are necessary to ensure that laboratories can properly diagnose TB. In addition, laboratories should have sufficient space to ensure the quality, safety and efficiency of the services provided to the clients whose samples are tested and to ensure the safety of laboratory personnel, patients and visitors.(5) Some countries experience difficulties with the implementation and quality assurance of sputum culture, which impacts upon this recommendation as it is dependent on access to quality assured laboratories that can offer TB culture. Sputum smear and culture examinations are also dependent on the quality of the sputum produced, so care should be taken to obtain adequate specimens and transporting them to the laboratory according to standard procedures to maintain viability of the bacilli to get a valid culture result.

In programmatic settings the practitioner treating MDR-TB patients is typically guided not only by bacteriological tests but also by markers of response to treatment or of disease progression, such as the patient's general condition, weight gain over time, resolution of disease manifestations, blood indices and the results of imaging (e.g. chest radiography).

## Monitoring and evaluation

Culture and microscopy results for tests performed in patients on MDR-TB treatment should be captured on the *Second-line TB treatment register* as well as the respective laboratory registers. (6) Sometimes these registers may exist as part of an electronic laboratory or patient information system, which facilitates greatly the access of data in real time by multiple users and can also help limit errors. It is important for the programme manager to assess the records on the *Second-line TB treatment register* for completeness of testing using both culture and sputum smear microscopy, any discordance between the two modalities, and whether decisions on regimen changes or assignment of outcome are coherent (e.g. does a case have sufficient negative culture test results available to classify as *Cured*?). Performance indicators such as contamination rates, turnaround times and proportion of culture tests done without results recorded in the patient information systems help improve quality of care.

## **Research priorities**

- Future analysis on predictors of treatment failure (related to strain, regimen and host), in addition to bacteriological response, in the following important subgroups would be helpful to identify more resource-saving and reduce time needed to make decisions
  - o Patients <15 years of age
  - o Extrapulmonary disease (different forms)
  - o Patients on shorter MDR-TB regimens (standardised or all-oral variants)
- It will also be helpful to keep assessing the potential role of future-generation rapid molecular testing beyond diagnostic testing, to monitor also treatment response
- Evaluation of engineering challenges to implement more affordable liquid culture systems

## **REFERENCES SUMMARY**

- 1. Guidelines for the programmatic management of drug-resistant tuberculosis, 2011 Update. (WHO/HTM/TB/2011.6) [Internet]. Geneva, World Health Organization. 2011. Available from: http://whqlibdoc.who.int/publications/2011/9789241501583\_eng.pdf
- 2. Kurbatova EV, Gammino VM, Bayona J, Becerra M, Danilovitz M, Falzon D, et al. Frequency and type of microbiological monitoring of multidrug-resistant tuberculosis treatment. Int J Tuberc Lung Dis. 2011 Nov;15(11):1553–5, i.
- 3. Mitnick CD, White RA, Lu C, Rodriguez CA, Bayona J, Becerra MC, et al. Multidrug-resistant tuberculosis treatment failure detection depends on monitoring interval and microbiological method. Eur Respir J. 2016 Oct;48(4):1160–70.
- 4. Implementing tuberculosis diagnostics: A policy framework (WHO/HTM/TB/2015.11) [Internet]. Geneva, World Health Organization. 2015. Available from: http://apps.who.int/iris/bitstream/10665/162712/1/9789241508612\_eng.pdf
- 5. Tuberculosis laboratory biosafety manual (WHO/HTM/ TB/2012.11) [Internet]. Geneva, World Health Organization. 2012. Available from: http://apps.who.int/iris/bitstream/10665/77949/1/9789241504638\_eng.pdf
- 6. Definitions and reporting framework for tuberculosis 2013 revision (WHO/HTM/TB/2013.2) [Internet]. Geneva, World Health Organization; 2013. Available from: http://apps.who.int/iris/bitstream/10665/79199/1/9789241505345\_eng.pdf

# Annex 10. Summaries of unpublished data and analysis plans used for the recommendations

Analysis plan to compare the outcomes of shorter MDR-TB treatment regimens with longer MDR-TB regimens using pooled individual patient data (IPD) – 2018 update

Investigators (incomplete list. Some yet to be confirmed):

F. Ahmad Khan, D. Menzies, D. Falzon, P. Zhi Li, A. Benedetti, V. Schwoebel, A. Trébucq, M.A. Hamid Salim, E. Casas, W. Sikhondze, P. du Cros, A. Piubello, MB Souleymane, M. Assao Neino, A. Van Deun, KJM Aung, J. Noeske, C. Kuaban, PIs from 9-country study, others to be added

### 1. Background

In 2016, WHO recommended the use of a standardized shorter regimen, of 9 to 11 months in duration, for the treatment of MDR-TB in patients without prior exposure to second-line drugs (SLD) and in whom resistance to fluoroquinolones (FQ) and second-line injectable agents (SLI) is unlikely.¹ The recommendation was based on aggregate and individual-patient data (IPD) meta-analyses of observational studies of standardized regimens of durations ranging from 9 to 14 months, all of which were based on the Bangladesh regimen (from hereafter, 'shorter regimens').¹,² At the time, unadjusted comparisons between outcomes from shorter regimens and those from regimens of more conventional composition and duration (from hereafter, 'longer regimens') were made using aggregated data, to inform the conclusions for the guidelines; however, the longer regimens data originated from an older IPD and were quite heterogeneous with respect to the era they were conducted, patient characteristics, and approaches to treatment, with studies using a variety of regimens including some with drugs that are now considered less effective.³

Since the 2016 guidelines meeting, IPD from an updated systematic review<sup>4</sup> of more contemporary studies of longer regimens have been compiled by the McGill University group. Ahead of a revision of its MDR-TB treatment policy in mid 2018, WHO has commissioned a new IPD meta-analysis using data from these studies, supplemented by data from cohorts of patients treated with either longer or shorter regimens identified via a public call for data issued by WHO in February 2018 (http://www.who.int/tb/features\_archive/public\_call\_treatment\_RR\_MDR\_TB/en/). WHO has requested McGill University to undertake an updated IPD meta-analysis that will include a comparison of outcomes between longer and shorter regimens using the standardised approach for development of evidence-based guidance and pursuant to a specific research question (http://apps.who.int/medicinedocs/documents/s22083en/s22083en.pdf). The remainder of this document describes a proposed approach for this analysis.

Two important points must be addressed in an analysis comparing shorter and longer regimens:

First, comparisons must account for differences in patient selection, and in key patient characteristics, between studies of shorter and longer regimens. In general, studies of the shorter regimens have excluded patients with, amongst other criteria, prior exposure to SLD, and have also excluded (or included very few, usually inadvertently) MDR-TB patients with additional resistance to the FQ or SLI used in the shorter regimen, or to both (i.e. XDR-TB).

Secondly, later generation fluoroquinolones (Gatifloxacin, Moxifloxacin, Levofloxacin) and SLIs are now widely considered key drugs for MDR-TB but have not always been used in patients treated with longer regimens, whereas the shorter regimen uses SLIs and Moxifloxacin or Gatifloxacin. Moreover, the composition and duration of longer regimens can vary in other important ways, between and within studies. The analytic approach will seek to minimize heterogeneity in key treatment approaches amongst the longer regimen studies, for example by excluding patients treated with Ofloxacin or Ciprofloxacin, and those not treated with a SLI (additional details described below).

### 2. Individual Patient Data for shorter and longer regimens —Sources

2.1. Shorter regimen studies included in the original IPD meta-analysis: (1) Bangladesh cohort published in Van Deun et al.<sup>5</sup> & Aung et al.<sup>6</sup> up until 2011 – investigators will provide updated data on additional patients since then; (2) MSF Uzbekistan<sup>2</sup> study interim data – to be updated with final data; (3) MSF Swaziland<sup>2</sup> study interim data — to be updated with final data.

- 2.2. Studies of the shorter regimens that were previously reported for the 2016 WHO guidelines revision only in aggregated format: (1) the 9 country study of the Union (Trebucq et al<sup>7</sup>); (2) previously published studies in Cameroon<sup>8</sup> and Niger <sup>9</sup> that were not in prior IPD;
- 2.3. New eligible studies identified through a call for data by the WHO (list to be completed by end April 2018).
  - 2.3.1. IPD from STREAM Stage 1 trial will not be available in time for this analysis. Tabulated results from this stage of the trial are expected to be combined with the IPD in an aggregate data meta-analysis (see Section 12.4).
- 2.4. Studies of the longer regimens, the majority of which were identified through a systematic review<sup>4</sup>, and through a public call issued by WHO in 2018.

*Note*: An updated systematic search of published literature will also be conducted to identify other studies of the shorter regimens, and also those of longer regimens, that have been published but where IPD data have not been provided to us. The overall results from identified studies will be used to assess if their exclusion from an IPD meta-analysis could have affected the results of the analyses described in this proposal.

### 3. PICO Question

The analysis proposed in this plan aims to answer the PICO (**P**opulation, Intervention, **C**omparator, **O**utcomes) question developed by the Guideline Development Group (GDG) appointed by WHO to review the latest evidence for the revision of its MDR-TB treatment policy. The wording of the PICO question and details on the four elements of the question (Table 1) are shown below:

In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) is a shorter treatment regimen (9-12 months) more or less likely to safely improve outcomes than longer regimens conforming to WHO guidelines?

Table 1.

Table 1.			
Population	Intervention	Comparator	Outcomes
MDR/RR-TB patients	Standardized	Longer regimens as	- Culture conversion by
NOT previously treated with	shorter MDR-TB	recommended by WHO	6 months
2nd line drugs, and with strains	regimen	defined as:	- Successful completion
susceptible to FQ and SLI	defined as:	- Use of at least 5 effective	of treatment
a. with severe disease	- duration of 9-12	drugs	- Bacteriological cure by
(cavities on radiography)	months	- Injectable agent given for	end of treatment
b. with additional drug resistance	- injectable agent	about 8 months, at least 4	- Treatment failure or
patterns for FLDs (PZA, INH, EMB) and ETO/PTO;	for 4-6 months	months after culture	relapse
c. with history of patient use	- combination of	conversion	- Survival (or death)
ethambutol, or pyrazinamide	drugs (usually 7 in	- Total treatment for about 20	- Adherence to
e. children (0-14y) / adults	the intensive phase	months, and at least 18	treatment (or treatment
(adolescents 10-19y if available);	and 4-5 in the	months past the date of	interruption due to non-
f. persons with HIV (+/- ARVs);	continuation)*	culture conversion to negative	adherence)
g. pregnant women;		- Injectable agent given until	- Adverse reactions
h. people with diabetes;		smear conversion and total	from anti-TB medicines
i. extrapulmonary disease j. malnutrition		treatment for at least 12	- Acquisition
j. mamutition		months after smear	(amplification) of drug
		conversion	resistance

<sup>\*</sup>Typically: [4-6]KmMfxPtoCfzZHhE/ [5 or 8]MfxCfzZE; Regimens utilizing Gfx or Lfx instead of Mfx; Eto instead of Pto; and Am or Cm instead of Km are also included. In some series Pto was included in the continuation phase. Regimens where total duration could reach 14 months due to treatment extension will also be included, as long as composition unchanged and the original intended duration was between 9-12 months.

This will be premised upon the patient subgroups in whom the shorter MDR-TB regimen may be recommended. Of note—patients with XDR and pre-XDR-TB will thus be excluded from both shorter and longer regimens for this analysis.

The restriction of the comparator group to patients with at least 5 effective drugs (see Section 6.1.1. for definition) implies that this PICO is asking how does the shorter regimen compare to longer regimens that comply with WHO recommendations about composition. Because the shorter regimen studies included have observed eligibility criteria and used a strictly standardized regimen, we will include patients treated with shorter regimens regardless of considerations about the number of effective drugs received.

### 4. Inclusion / Exclusion Criteria

Note: for all DST related criteria, if both phenotypic and genotypic results are available, we will use results from the former. Only Xpert, LPA, or sequencing will be accepted as methods of genotypic DST.

### 4.1. <u>Inclusion criteria</u>

MDR-TB: DST-confirmed resistance to Inh and Rif

RR-TB: DST-confirmed resistance to Rif.

No prior exposure to  $2^{nd}$  line drugs:  $\leq 30$  days of treatment with SLD.

### 4.2. Exclusion criteria

DST-confirmed susceptibility to Inh (Section 12.1 describes a sensitivity analysis where these are included.)

DST-confirmed XDR-TB or pre-XDR-TB

Treatment regimen that does not contain a SLI or either Moxifloxacin, Gatifloxacin, or Levofloxacin (high-dose) Missing information on duration of treatment

4.2.1. Note re second-line injectable resistance: In two studies of shorter regimens, the protocol called for MDR-TB with resistance to a SLI to be treated with an alternative SLI to which the strain had demonstrated baseline susceptibility. Patients in this situation—with baseline resistance to one SLI but treated with a SLI to which there infecting strain was susceptible—will be included in the IPD.

### 5. Intervention regimens

5.1. Definition of shorter regimens – see also Section 3 and Table 1. We will operationalize this as follows: Typically: [4-6]KmMfxPtoCfzZHhE/ [5 or 8]MfxCfzZE; Regimens utilizing Gfx or Lfx instead of Mfx; Eto instead of Pto; and Am or Cm instead of Km are also included. In some series Pto was included in the continuation phase. Regimens where total duration could reach 14 months due to treatment extension will also be included, as long as composition unchanged and the original intended duration was between 9-12 months.

As the shorter regimen recommended by WHO is strictly standardized, patients treated with shorter regimens whose composition deviated from the above will be excluded in our analyses.

### 6. Comparator regimens

- 6.1. <u>Definitions of Longer regimens</u>: see also Section 3 and Table 1. The parameters defining the comparator are based on WHO recommendations for DR-TB treatment.\* Per WHO, longer MDR-TB regimens are treatments for RR-TB or MDR-TB which last 18 months or more and which may be *standardized or individualized*. These regimens are usually designed to include a minimum number of second-line TB medicines considered to be effective based on patient history or drug-resistance patterns. These regimens were previously qualified as "conventional", having been the mainstay of MDR-TB treatment before the 2016 WHO guidelines update.
  - 6.1.1. *Number of effective drugs:* In line with the PICO question the comparator group will be restricted to patients treated with at least 5 effective drugs. This follows the WHO

<sup>\*</sup> The characteristics of current longer and shorter regimens are described in the WHO treatment guidelines for drug-resistant tuberculosis, 2016 update (apps.who.int/iris/bitstream/10665/250125/1/9789241549639-eng.pdf). The previous WHO recommendation for a longer regimen can be found in the 2011 edition of the guidelines (whqlibdoc.who.int/publications/2011/9789241501583\_eng.pdf). WHO-recommended longer regimens referred to in this Annex have a duration of 18 months or more.

recommendation to use a minimum of 4 SLDs plus pyrazinamide, replacing any of these 5 if they are considered not effective. No upper limit on the number of effective drugs in a longer regimen will apply. When selecting comparator patients on the longer regimen IPD, a drug will be classified as effective if it was used in a regimen and either DST demonstrated the patient's infecting strain was susceptible, or, DST had not been performed on an individual's strain AND:

- 1- DST data were available for  $\geq$ 50% of patients from the same cohort amongst whom prevalence of resistance to the drug was  $\leq$  10%, or,
- 2- DST data were available for <50% of the patient's cohort, and DST data were available for  $\ge 50\%$  of patients from the same country amongst whom prevalence of resistance was < 10%, or,
- 3- DST data were available for <50% of patients from the same cohort, and also available for <50% of patients from the same country, and the prevalence of resistance to the drug was <10% amongst all patients in the IPD dataset included in the analysis for this PICO question.

While DST is uncommonly performed for bedaquiline and delamanid, because primary resistance to these medications is expected to be uncommon in patients not previously treated with second-line drugs, we will count these as effective drugs unless DST demonstrated resistance.

6.1.2. *Duration*: (a) Total Duration: We will include patients in the longer regimens group if they were treated for ≥ 18 months. Patients treated for <18 months will be included if: (1) their outcome was Failure, Lost to follow up, or Died. It is important to include these groups treated for < 18 months to avoid bias by reverse causation (when an outcome is assigned prior to reaching the intended duration of treatment, the outcome becomes a determinant of duration, rather than the other way around). Patients treated for <18 months who were assigned an outcome of Success will be excluded from analyses (see Section 6.2.3 for details).

Rationale for using 18 months: While, the PICO specifies that comparator regimens should have an overall treatment duration of 20 months, data from studies of longer regimens suggest that the intended duration in many centres was less than 20 months: amongst successfully treated patients, the IQR for total duration was [18.7-24 months], and median total duration was 21.5 months.

(b) Intensive phase duration: We will include patients in the longer regimens regardless of the duration of intensive phase treatment.

Rationale: We will not place a lower limit on the duration of the intensive phase for a patient to be considered successfully treated because (a) patients in whom the injectable was stopped early could still meet criteria for Success by Laserson and WHO 2013 criteria; and (b) 42% of successfully treated patients in studies of longer regimens had an intensive phase of < 8 months in duration, suggesting this was considered adequate in many centres.

6.1.3. Regimen composition: Patients treated with agents from those recommended by WHO for use in longer regimens (see Table 6 of the 2016 guidelines<sup>1</sup>) will be included (provided other selection criteria are met, and no exclusion criteria are present). This means patients treated with longer regimens with newer agents like bedaquiline, delamanid, and linezolid will also be eligible for inclusion in the longer regimens group.

### 6.2. Exclusions from Comparator Group

- 6.2.1. Patients treated without levofloxacin, moxifloxacin or gatifloxacin, or without a SLI.
- 6.2.2. Patients treated with less than 5 effective drugs.
- 6.2.3. Patients with treatment duration of less than 18 months if assigned an outcome of Success

- 6.2.3.1. Rationale for exclusion: Nearly all studies of longer regimens used either Laserson/WHO 2013 definitions for treatment outcomes, hence patients should not have been assigned an outcome of Success if they were treated for less than the intended duration. So if we assume that the intended duration of treatment was 18 months, then technically patients treated for less than this should not have been declared Successfully treated. However, in the longer regimens data, ~15% of successfully treated patients were treated for <18 months. It is possible that the inconsistency between the assigned outcome of Success in this subset of patients, and the stated use of Laserson/WHO 2013 criteria, signals an error in data on duration, or data on the outcome, or, that some centers considered < 18 months of treatment acceptable for some patients (amongst those reported as successfully treated after <18 months of therapy, ~46% had been treated for 16-17.9). We judged that reclassification of these patients' outcomes as Lost to follow-up would be inappropriate when taking into consideration that most centers used individualized treatment, and reported using Laseron/WHO 2013 outcome definitions. On the other hand, we judged that it would be inappropriate to include in the longer regimens group patients assigned an outcome of Success despite being treated for <18 months, when our intention is to compare longer and shorter regimens as formulated by the PICO. We judge that the least biased way to handle this inconsistency is to exclude this subset of patients from analyses. See Table 2.
- 6.2.4. Patients who died 12 months after starting treatment will be excluded from comparisons between longer and shorter regimens See **Table 2**.
  - 6.2.4.1. Rationale for exclusion: Our outcome of death is restricted to deaths that occur while on MDR treatment. As such, with a shorter treatment duration there is less time during which deaths can accrue as compared to with a longer duration. In the shorter regimen studies included in the IPD, treatment lasts up to 12 months. Hence to avoid bias due to differential ascertainment of death between shorter and longer regimens, we will exclude patients that died 12 months after starting treatment.

Table 2. Handling of data from the longer regimens group based on treatment duration and outcome assigned in their original study.

Actual duration of treatment with a longer MDR-TB regimen, in months						
Original outcome	< 18	18-24	>24			
Success	Exclude <sup>1</sup>	Include	Include <sup>2</sup>			
Failure or Relapse	Include	Include	Include			
Died during treatment	Include if death in first 12 months <sup>3</sup>	Include if death in first 12 months <sup>3</sup>	Include if death in first 12 months <sup>3</sup>			
Loss to follow-up (LFU)	Include	Include	Include <sup>2</sup>			

<sup>&</sup>lt;sup>1</sup> See Section 6.2.3.1. for explanation. <sup>2</sup> See Section 12.2 for explanation. <sup>3</sup> See Section 6.2.4.1. for explanation.

7. Outcomes -- We will keep the outcomes assigned in the original studies and no reclassification of outcomes will be done. The following observations support this position: (1) nearly all studies have used either WHO 2013<sup>10</sup> or Laserson<sup>11</sup> criteria to assign outcomes, (2) the previous analysis undertaken to inform the 2016 WHO recommendation on the use of a shorter MDR-TB regimen has adhered to this approach, as well as other publications from the longer regimen IPD group (in agreement with the Collaborative IPD group<sup>3,12</sup>), and (3) we have not been provided with all the data required to re-classify patient outcomes through independent verification of data from all studies.

### 7.1. <u>Definitions of treatment outcomes:</u>

Success: cure or treatment completion, as defined in the original studies.

Failure: For shorter regimen studies, we will use the definition used in the original studies as long as they are similar to definitions used in the published studies. The majority of longer regimen studies has defined outcomes based either on the Laserson definitions or the 2013 WHO criteria. Relapse: Will be defined per WHO, with patients being classified as Relapse if they had "previously been treated for TB, were declared cured or treatment completed at the end of their most recent course of treatment, and are now diagnosed with a recurrent episode of TB (either a true relapse or a new episode of TB caused by reinfection)".

*Death*: We will only count deaths during treatment, excluding deaths for patients who die after month 12 of treatment (See 6.2.4 for explanation).

Lost to follow-up: This will be defined as interruption of treatment for  $\geq 2$  consecutive months without medical approval and not meeting criteria for failure. These criteria are applied by shorter regimen studies and are similar to those used in most longer regimen studies and to the WHO definitions in 2005 and 2013.

## 7.2. <u>Analyses of treatment outcomes</u>

The outcomes selected for inclusion in the GRADE evidence summary† tables will be selected from amongst the ones in the PICO questions (all of these outcomes have been scored by the experts of the WHO Guideline Development Group (GDG) for the 2018 MDR-TB treatment guidelines as being "important" or "critical").

Previous analysis done for the WHO 2016 guidance have focused on the following mainstay outcomes:

Analysis 1: Success versus Failure/Relapse (excluding Death & Loss to Follow Up) Analysis 2: Success versus Failure/Relapse/Death (excluding Loss to Follow Up) Analysis 3: Success versus Failure/Relapse/Death/Loss to Follow Up

Similar analyses for other outcomes may be expected as requested by the GDG (see others in PICO table): Culture conversion at 6 months; Adverse events; Acquired Drug Resistance.

- 7.2.1. Note: If requested by GDG, the analyses maybe re-formulated as follows:

  Death versus Success/Failure/Relapse
  Failure/Relapse versus Success
- **8.** Covariates. The following are considered key covariates to be used for stratified analyses, or adjusted for in multivariable analyses: Age, sex, HIV, ART if HIV, AFB, Cavitation on CXR, Prior treatment with first line drugs.
  - 8.1. Multivariable meta-regression analyses will adjust for each of the above. For multivariable analyses only, we will impute missing values for these key covariates. Individuals will be assigned within-study means for values of covariates for which their information is missing. In the absence of within-study means (e.g. all participants from the same study are missing data on the covariate), we will use the mean value of all patients included in the IPD and treated with the same regimen type (i.e. shorter or longer). Note that we are not interested in estimating the effect of these variables on treatment outcomes, we are only interested in adjusting for potential confounding due to differences in these covariates, hence our use of mean values to impute missing data.
    - Rationale: Because we are adjusting for many covariates, excluding patients with missing data could result in a substantial proportion being excluded from analyses (e.g. if for each of these 7 covariates, 5% of patients are missing data, then 35% of all patients will end up excluded).
  - 8.2. <u>FQ type and dose</u> When possible, we will try to adjust for FQ type and dose (e.g. limiting to longer regimen patients receiving Mfx 400mg/d). It may not be possible due to small sample size.

\_

<sup>†</sup> GRADE homepage. www.gradeworkinggroup.org

- 9. Pooled Percentages of Each Treatment Outcomes. We will use random-effects aggregate data meta-analysis to calculate pooled percentages of each treatment outcomes (i.e. where numerators are Success, Failure/Relapse, Death, or Loss to follow-up, with the denominator the sum of all outcomes), pooling all shorter regimen studies, and separately pooling longer regimen studies. These analyses will not be matched or adjusted. For pooled percentages, heterogeneity will be estimated through the variance of the random-effects parameter, and considered significant if its confidence interval does not include 0.2
  - 9.1. For Death, when reporting the pooled percentage: in the longer regimen group, we will report the percentage excluding deaths that occurred after 12 months of treatment (See Section 6.2.4.1 for rationale of avoiding ascertainment bias). In a footnote, we will also report the proportion that died counting all deaths regardless of their timing.
- 10. Analyses: Adjusted Odds Ratios and Absolute Risk Differences –We plan to do multivariable IPD meta-regression (either propensity score matched, or through standard multivariable models) to estimate adjusted odds ratios and adjusted risk differences, comparing shorter regimens to the longer comparator.
  - 10.1. Comparison: Shorter regimens versus Longer regimens
    - 10.1.1. <u>Aggregate data meta-analyses</u> (unmatched, random effects meta-analysis) will be performed to estimate the pooled percentage (i.e. risk) of each outcome:
      - 10.1.1.1. Including all patients
      - 10.1.1.2. Stratified by baseline resistance to the following drugs: PZA; EMB; ETO/PTO. Rationale: Stratifying by baseline resistance to a drug will provide some information about which drugs/resistance profiles are determinants of the effectiveness of shorter regimens relative to longer regimens. If sample size of a strata of drugresistance is too small to perform meta-analysis, we will use simple (unweighted) pooling.
      - 10.1.1.3. Stratified by key patient subgroups identified in the PICO table.
    - 10.1.2. IPD meta-regression The "basic" model will compare shorter to longer regimens, and adjust for age, sex, HIV, ART if HIV, AFB, prior treatment with first line drugs (the "basic" model). Recall that patients with FQ-resistance or SLI-resistance are excluded from all analyses, and also that we have restricted the longer regimen group to those with at least 5 effective drugs. Because this basic model does not additionally adjust for baseline resistance, it addresses the question of how shorter and longer regimens compare when the latter are given under conditions where the minimum recommended number of effective drugs can be achieved. In addition to the basic model, we will explore sub-group effects either by including interaction terms or through stratified analyses. The following sub-group effects will be explored:
      - 10.1.2.1. Baseline resistance to PZA, EMB, ETO/PTO, adjusting for all covariates in the basic model. These models address questions about whether resistance to a drug at baseline affects the relative effectiveness of the shorter regimen compared to the longer regimen, when the latter is given under conditions where the minimum recommended number of effective drugs can be achieved.
      - 10.1.2.2. HIV status, adjusting for all covariates in the basic model. This addresses the question of whether HIV affects the relative effectiveness of the shorter regimen compared to the longer regimen, when the latter is given under conditions where the minimum recommended number of effective drugs can be achieved.
      - 10.1.2.3. Severe disease defined as the presence of cavities *or* bilateral disease on chest radiography, or smear-positivity.
      - 10.1.2.4. Children (0-14), Diabetes, Extrapulmonary disease, and Pregnancy. However it is unlikely that we will have sufficient patients in these categories to undertake these analyses.

In addition to the above

- 10.1.2.5. For IPD meta-regression, we will use generalized linear mixed models (SAS PROC GLIMMIX). Approaches to modelling of random-effects, propensity-score matching, and estimation of heterogeneity, will follow those used by the McGill group for other PICO questions informing the 2018 WHO MDR/RR-TB Guidelines, and the reader is referred to those analysis plans for methodological details. Generally, we will adjust for confounders using either propensity-score matching or multivariable meta-regression, and we will use random-effects models unless these present computational challenges (e.g. non-convergence), in which case fixed-effects models will be used.
- 11. Analyses for Other Outcomes- For the following outcomes, if sufficient data are available, IPD meta-analyses will be undertaken. Analyses will be restricted to aggregate data meta-analysis if only aggregate data are available, and we will use simple (i.e. unweighted) pooling if meta-analyses are not possible.
  - 11.1. <u>Culture conversion by 6 months</u>—the outcome will be dichotomized (converted by month 6 versus not). Data available are variable amongst the studies. For studies that provide dates of cultures, we will use a cut-off of 182 days from treatment initiation to define the cut-off of 6 months.
  - 11.2. <u>Adverse events (AE)</u>— If possible, we will calculate pooled percentages, overall and stratified by organ system and severity. Data may be quite limited. If pursuing the analysis, we will dichotomize this variable, defining the outcome as experiencing any AE of severity ≥ 3, or as drug stopped for AE. If aggregate meta-analysis does not work, we will use simple (i.e. unweighted) pooling.
  - 11.3. <u>Acquired Drug Resistance</u>—the outcome will be dichotomized, defined as acquisition of drugresistance to any medication for which susceptibility was confirmed at baseline. Data may be quite limited. If aggregate meta-analysis does not work, we will try simple pooling.
  - 11.4. Effect of type and dose of FQ on treatment outcomes amongst patients treated with shorter regimens This is not a specific PICO question, but has been identified as being of interest by members of the GDG. The analysis will be an IPD metaregression, restricted to patients treated with the shorter regimen, and will estimate the effect of type of quinolone on treatment outcomes, in model adjusting for the covariates described in the Basic model (Section 8). Usual-dose Mfx, high-dose Lfx, high-dose Mfx, usual-dose Gfx, will each be compared to high dose Gfx.

### 12. Sensitivity analyses

- 12.1. Including patients with Inh-susceptible disease While most studies of shorter regimens have included patients with Rif-R TB regardless of Inh DST results, our analyses will exclude patients with Rif-R Inh-S longer regimen studies have been restricted to MDR defined as Rif-R + Inh-R. To assess the impact of excluding patients with Inh-Susceptible disease, we will perform this sensitivity analysis. We expect that the large majority of patients would have documented resistance to Inh and therefore this analysis will add only a small proportion of patients. Nonetheless this is an important analysis to make given that WHO recommends an MDR-TB treatment regimen for all Rif-R cases regardless of INH susceptibility, and the PICO is inclusive of all MDR/RR-TB cases.
- 12.2. Sensitivity analyses excluding all patients treated for > 24 months
  - 12.2.1. Rationale: Usually, the intended duration of treatment with a longer regimen is between 18 to 24 months. However, our dataset includes patients treated with longer regimens for more than 24 months, some of whom were assigned an outcome of Success. There could be an argument that the inclusion of these cases in the analysis could bias against the longer regimen (e.g. inclusion of cases with unrecognised drug resistance, extensive disease, or compromised adherence) or in favour (e.g. important changes made to the regimen without assigning "Treatment failure" as an outcome). This sensitivity analysis is being undertake to assess this.

Note: Early feedback on this proposal included suggestions that all patients treated for > 24 months have their outcomes classified as Failure. However, 51 of 54 studies of longer regimens used either WHO 2013 or Laserson criteria for outcomes-- so the original studies would have already assigned an outcome of Failure if the treatment regimen was terminated early (per Laserson) or drugs changed (WHO 2013). As explained in Section 7, we will not be re-classifying outcomes because we do not have the data to do so, and re-classification without original data is likely less accurate and may introduce information bias as compared to accepting the outcomes assigned by original investigators using standardized criteria that are known to TB programmes worldwide. Moreover, in line with other analyses conducted to inform the WHO DR-TB treatment guidelines, no patients are having their outcomes re-classified—hence, doing so just for the analysis comparing shorter to longer regimens is inconsistent with how longer regimen data will be handled for the majority of the guidelines, and also with how these data have been used to inform the guidelines of other professional bodies (ATS/CDC/IDSA/ERS).

- 12.3. Subgroup analyses restricted to participants with pre-XDR. Depending on the numbers of patients available, we will attempt to compare longer and shorter regimens in this subgroup (which are otherwise excluded from the IPD as the WHO recommendation is to not use the shorter regimen in these circumstances).
  - 12.3.1. Rationale: These patients are not usually eligible for the shorter regimen but have at times been inadvertently included and later allowed to complete the regimen after a belated diagnosis.
- 12.4. Aggregate data meta-analyses including aggregate results from the STREAM Stage 1 –IPD data from the STREAM Stage 1 trial will not be available in time for the WHO GDG in July 2018. However, aggregate data from the preliminary results of this stage of the trial have already been communicated and an update is expected in time for the July meeting. We will combine these with aggregate outcome data with IPD from the observational studies. This analysis will report pooled percentages of each outcome, using traditional random-effects meta-analysis, stratified by shorter versus longer regimens. If feasible, an aggregate meta-regression will be performed, providing an unadjusted odds ratio comparing shorter and longer regimens.

# Number of patients included in main & sensitivity analyses for each shorter regimen study

Study	In main analysis	In sensitivity analyses†	Total
1-Bangladesh	965	124	1089
2-Uzbekistan MSF	122	5	127
3-Eswatini MSF	114	11	125
4-Cameroon	415	10	425
5-Niger	162	20	182
6-Union 9 Country	867	137	1004
7-Tajikistan KNCV	16	1	17
8-Kyrgyzstan KNCV	15	1	16
9-South Africa	45	137	182
TOTAL	2721	446	3167

†Inh-susceptible, FQ resistant, SLI resistant, or modified standardized shorter regimen (in which drugs that are not part of the usual regimen were added)

## References

- 1. World Health Organization. WHO treatment guidelines for drug-resistant tuberculosis 2016 update. Geneva: World Health Organization; 2016.
- 2. Ahmad Khan F, Salim MAH, du Cros P, et al. Effectiveness and safety of standardised shorter regimens for multidrug-resistant tuberculosis: individual patient data and aggregate data meta-analyses. *Eur Respir J* 2017; **50**(1).
- 3. Ahuja SD, Ashkin D, Avendano M, et al. Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: an individual patient data meta-analysis of 9,153 patients. *PLoS Med* 2012; **9**(8): e1001300.
- 4. Bastos M, et al. An Updated Systematic Review and Meta-Analysis for Treatment of Multidrug-Resistant Tuberculosis. *European Respiratory Journal, requested revision under consideration* 2016.
- 5. Van Deun A, Maug AK, Hamid Salim MA, et al. Short, highly effective, and inexpensive standardized treatment of multidrug-resistant tuberculosis. *Am J Resp Crit Care Med* 2010; **182**(5): 684-92.
- 6. Aung KJ, Van Deun A, Declercq E, et al. Successful '9-month Bangladesh regimen' for multidrug-resistant tuberculosis among over 500 consecutive patients. *Int J Tuberc Lung Dis* 2014; **18**(10): 1180-7.
- 7. Trebucq A, Schwoebel V, Kashongwe Z, et al. Treatment outcome with a short multidrug-resistant tuberculosis regimen in nine African countries. *Int J Tuberc Lung Dis* 2018; **22**(1): 17-25.
- 8. Kuaban C, Noeske J, Rieder HL, Ait-Khaled N, Abena Foe JL, Trebucq A. High effectiveness of a 12-month regimen for MDR-TB patients in Cameroon. *Int J Tuberc Lung Dis* 2015; **19**(5): 517-24.
- 9. Piubello A, Harouna SH, Souleymane MB, et al. High cure rate with standardised short-course multidrugresistant tuberculosis treatment in Niger: no relapses. *Int J Tuberc Lung Dis* 2014; **18**(10): 1188-94.
- 10. World Health Organization. Definitions and reporting framework for tuberculosis 2013 revision. WHO/HTM/TB/2013.2. Geneva; 2013.
- 11. Laserson KF, Thorpe LE, Leimane V, et al. Speaking the same language: treatment outcome definitions for multidrug-resistant tuberculosis. *Int J Tuberc Lung Dis* 2005; **9**(6): 640-5.
- 12. Falzon D, Gandhi N, Migliori G, et al. Resistance to fluoroquinolones and second-line injectable drugs: impact on MDR-TB outcomes. *Eur Respir J* 2012.

## Determinants of treatment outcomes of multidrug-resistant tuberculosis (MDR-TB) Statistical Analysis Plan for an individual patient data meta-analysis

#### **Abbreviations**

WHO World Health Organization

IPD-MA Individual Participant Data Meta-Analysis

IQR Interquartile Range

TB Tuberculosis

MDR-TB Multidrug Resistant Tuberculosis
XDR-TB Extensively Drug Resistant Tuberculosis

AE Adverse Event
LTFU Loss to Follow-up
aOR Adjusted Odds Ratio
aRD Adjusted Risk Difference

LPA Line Probe Assay

DST Drug Susceptibility Test

AFB Acid-fast Bacilli

ROBINS-I Risk of Bias in Non-Randomized Studies of Interventions

FLD First-line Drugs
SLD Second-line Drugs
PZA Pyrazinamide
RIF Rifampicin
FQ Fluoroquinolone
SLI Second-line Injectable

Eto Ethionamide Prothionamide Pto Cs Cycloserine Terizidone Trd Lzd Linezolid Cfz Clofazimine Kanamycin Km **DLM** Delamanid Bedaquiline **BDQ** 

#### **Background**

- Current treatment outcomes of MDR-TB are suboptimal and many of the drugs utilized confer high risk of toxicity to patients.
- MDR-TB treatment generally consists of an initial intensive phase where an injectable agent (e.g. amikacin) is utilized, followed by a continuation phase where injectable use is stopped.
- Treatment effectiveness is monitored using sputum smear microscopy and/or sputum culture and drug selection is generally guided by drug susceptibility testing (LPA, phenotypic).
- With various options for treatment (including new drugs like bedaquiline and delamanid) and treatment
  monitoring, it is imperative to determine if use of specific drugs, use of drugs for specific durations, the
  number of drugs used, and/or use of specific monitoring strategies improve patient outcomes.

#### ABOUT THE 2018 WHO MDR/-RR-TB GUIDANCE REVISION

The WHO treatment guidelines for rifampicin- and multidrug-resistant tuberculosis, 2018 update will address key questions of topical debate on which TB authorities in Member States and other implementers demand guidance from WHO. The scope of the new guidelines will cover key questions included in the 2011 and 2016 editions of the drug-resistant TB treatment guidance<sup>1,2</sup>, as well as other emerging topical areas relating to newer medicines. Questions worded in PICO format (Population, Intervention, Comparator, Outcome) were proposed by the WHO Guideline Development Group in May 2018 to guide the collection of evidence and formulation of recommendations:

- **Q1.** In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) is a shorter treatment regimen (9-12 months) more or less likely to safely improve outcomes than longer regimens conforming to WHO guidelines<sup>3</sup>?
- **Q2.** In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents are more likely to improve outcomes when forming part of a longer regimen conforming to WHO guidelines?
- **Q3**. In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines are outcomes safely improved with fewer or more than five effective medicines in the intensive phase?
- **Q4.** In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines are outcomes safely improved with an intensive phase shorter or longer than eight months?
- **Q5.** In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines are outcomes safely improved with a total duration shorter or longer than twenty months?
- **Q6.** In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) on longer regimens composed in accordance with WHO guidelines what is the minimum duration of treatment after culture conversion that is more likely to improve outcomes?
- **Q7.** In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB) treated with longer or shorter regimens composed in accordance with WHO guidelines, is monitoring using monthly cultures, in addition to smear microscopy, more likely to detect non-response to treatment?

This document proposes an analysis of pooled individual patient data (IPD) for MDR/RR-TB patients on treatment to address the PICO questions 2-7 (PICO 1 is addressed in a dedicated analysis plan separate from this document). Amongst other things, this analysis plan proposes an adjustment to the parameters of the original PICO questions based on the available data and explains the rationale for these changes.

The characteristics of current longer and shorter regimens are described in the 2011 and 2016 WHO treatment guidelines for drug-resistant tuberculosis

<sup>&</sup>lt;sup>1</sup>Guidelines for the programmatic management of drug-resistant tuberculosis, 2011 Update. (WHO/HTM/TB/2011.6). Geneva, World Health Organization. 2011. http://whqlibdoc.who.int/publications/2011/9789241501583\_eng.pdf

WHO treatment guidelines for drug-resistant tuberculosis, 2016 update (WHO/HTM/TB/2016.4). Geneva, World Health Organization. 2016. http://apps.who.int/iris/bitstream/10665/250125/1/9789241549639-eng.pdf

#### **Objectives**

(SAP for PICO1 – short regimens – is a separate document)

For specific populations to be analyzed in each of these objectives, please see Section 2 of the Methods.

PICO 2. In patients who receive  $\geq 18$  months of treatment and in patients who receive a minimum of 3 likely effective drugs  $\pm$  PZA (not counting the drug of interest so end result will be usually minimum of 4 effective drugs assuming the drug of interest is effective!) in the initial phase, to identify individual drugs that are more likely to improve outcomes of successful completion of treatment or bacteriological cure, acquisition of drug resistance, and culture conversion by 6 months.

- For the analysis of death, it is not possible to specify a minimum duration, because death (i.e. the outcome) determines the duration (the 'determinant' variable of interest). Hence this analysis will be restricted to patients treated at centers where ≥75% of patients with treatment success (cure or completion) receive at least 18 months of treatment. By limiting analysis to centers where the majority of successfully treated patients reach this treatment duration it can be reasonably assumed these patients would have otherwise received this treatment duration.
- For this analysis, patients ('controls') who didn't receive a drug with 'limited availability' (e.g. delamanid, bedaquiline) will be selected from data sets where that drug was not available, meaning that the drug was not used at all in that centre/site. This will reduce confounding by indication since these patients' characteristics could not, by definition, have affected their receiving this drug.
- For analysis of bedaquiline and delamanid, some data is from randomized controlled trials. For the analysis of each of these two specific drugs, analysis will be presented separately for outcomes measured in these trials from all other cohorts in the IPD. In the case of bedaquiline, for other drug-specific analyses and PICOs, or outcomes not measured in the RCTs, trials will be treated as cohorts within the IPD. In the case of delamanid, Trial 213 was only provided for use in this specific PICO question with respect to delamanid, therefore for other drug-specific analyses, it will be excluded.
- We have set a minimum of 3 likely effective drugs (<u>+</u> PZA) to be included with the drug of interest. Note that if the drug of interest is effective then this gives minimum 4 effective drugs, not counting PZA. In sensitivity analyses, we can examine increasing that minimum to 5, ignoring use of PZA, or even restrict to regimens that closely match WHO (2016) recommendations of 1 FQN, 1 SLI, 2 others (ideally from Group C), and PZA. This will be also be thoroughly described (how many patients in/excluded with different criteria for minimum regimens) and the effect on outcomes.

**PICO 3**. In patients who receive  $\geq$ 18 months of treatment, what is the number of likely effective medicines in the intensive (injectable) phase (reference: 5 effective medicines) that results in optimal outcomes of successful completion of treatment or bacteriological cure, acquisition of drug resistance, and culture conversion by 6 months relative to the reference group?

• For the analysis of death, it is not possible to specify a minimum duration, because death determines the duration (rather than the variable of interest determining the outcome). But this IPD2018 data set included only cohorts of patients treated at centers where the majority of patients (unless there are specific reasons to treat with less) receive at least 18 months of treatment. By limiting analysis to centers where the majority of successfully treated patients reach this treatment duration time it can be reasonably assumed these patients would have otherwise received this treatment duration.

**PICO 4-6**. In patients receiving a minimum of 4 effective drugs (<u>+</u>PZA) in the initial (injectable) phase and 3 effective drugs in the continuation phase, for each treatment interval listed below, which duration results in

optimal successful completion of treatment or bacteriological cure, acquisition of drug resistance, and culture conversion by 6 months, relative to the reference group?

**PICO 4**. Intensive (Injectable) Phase (reference: ≥7 and <8.5 months, Compare with longer and shorter)

PICO 5. Total Duration of Treatment (reference: ≥18 and <20 months, Compare with longer and shorter)

**PICO 6.** Duration of intensive (injectable) phase post-culture conversion (reference: ≥3 and <5 months. Compare longer and shorter). Total duration of treatment post-culture conversion (reference: ≥12 and <15 months. Compare with longer and shorter)

- The outcome of death cannot be analyzed for PICO 4-6. As noted above for PICO 2 and 3, duration analyses for the outcome of death are biased, as death determines treatment duration.
- Failure is our primary outcome in this analysis (de facto, since relapse is measured in few studies). The relationship of failure with duration is also potentially biased in an observational study. Failure may be declared early, and truncate therapy, or therapy may be prolonged because of apparent failure (patient not responding well). This problem should be kept in mind when interpreting results.
- For PICO 5, total treatment duration will be analyzed including all centres using the conventional regimen (defined operationally for this analysis as all regimens *that are not explicitly the 'short' or 'Bangladesh' regimen*.) A sensitivity analysis (see Methods, Section 6) is planned to quantify potential biases with this approach.
- For PICO 4 and 6, analysis will be restricted to patients treated at centers where ≥75% of patients with treatment success (cure or completion) received at least 18 months of treatment, for reasons mentioned in PICO 2 and 3.
- In PICO 6, there is potential bias stemming from change in clinical care depending on *timing* of culture conversion. Hence a sensitivity analysis is planned to determine outcomes associated with different durations, based on timing of culture conversion (see Methods, Section 6).

**PICO 7**: In individuals receiving a minimum of 4 effective drugs in the initial (injectable) phase and 3 effective drugs in the continuation phase, is the use of culture, in addition to smear microscopy, more likely to detect treatment failure? What is the predictive ability of failure to culture convert by month *x* for the outcome of treatment failure?

• The outcomes here would be defined as number failed/number successes in patients who did vs. did not convert by month x. Analysis would thus be comparison of proportions.

#### **Methods**

#### 1. 1. Study population

The study populations may vary depending on the specific objectives being analyzed, however they all conform to this patient definition:

- Patients with MDR-TB, including XDR-TB and RIF-resistant TB (RIF mono-resistant; cohorts entirely or nearly entirely (>80%) diagnosed with Gene Xpert without further DST will be excluded):
  - o a) MDR-TB confirmed by phenotypic or genotypic tests;
  - o b) had pulmonary TB, with or without concomitant extra-pulmonary involvement;
  - o c) with complete drug treatment information;

o d) had an end of treatment outcome as defined by Laserson/WHO (i.e., were not still on treatment when the dataset was finalized and sent to McGill).

Initial analyses for each objective will include the entire population described in the objective. Planned secondary analyses restricted to population subgroups are described in the sections on sensitivity analyses for each PICO question.

#### 1.2. Data sources

We have created a new IPD in MDR data-set – the *2018 IPDinMDR* database. We have sought agreement of all authors/investigators who are providing new patient data to add to our currently available *2016 IPDinMDR* database. We have added 625 patients from 8 other datasets. These include the following: Belarus (98 patients), Latvia (72 patients), one tertiary care centre in Brazil (100 patients), France (10 patients added to cohort included in 2016 IPD), South Korea (25 patients), Russia (125 patients), End-TB (166 patients), Australia (29 patients). We have also received data from Otsuka (compassionate delamanid use - data for 9 patients, data from Trials 204, 208, and 116 (421 patients) and Trial 213 (511 patients – see Section 1.3 below) Data sets with only 6 month outcomes were not included, but if 6 month outcomes and end of treatment outcomes are available in the same data set, this information will be used for PICO7.

After obtaining agreement from all new sources, we will map all variables to ensure that they have the same definition and are in the same format as the databases described below:

- 2016 IPD in MDR: 50 studies with 12,030 patients
- 2010 IPD in MDR [1]: 31 studies with 8,955 patients

From here, data from the 2016 IPD and all newly received data will be combined to form the **2018 IPDinMDR** dataset to answer the questions described in the aforementioned objectives.

Systematic reviews were completed to identify studies for the 2010 and 2016 IPDs. For this update, an updated systematic review has not been completed due to the limited time available. Instead a public call was made by the WHO to contribute data.

#### 1.2.1 South African dataset

Individual data for all patients treated for MDR-TB in public facilities in South Africa (SA) between 2010 to early 2016 were provided by the Department of Health of SA. Among the 56,347 patients however 21% were lost to follow up and more than half of these records had incomplete information on variables essential for the pooled IPD analysis (specifically DST to FQ and SLI, HIV status, and ARV treatment, age, sex, AFB smear, and prior treatment). Over 1,900 patients received bedaquiline, mostly in 2014-2015, and 6 received delamanid. From all 5482 patients who began MDR treatment between January 1, 2015 and December 31, 2015, we will include **all** those who received bedaquiline (N=1210) and a random sample of all those who did not receive BDQ – to equal twice as many patients (N=2420). Resistance patterns in the complete cohort of individuals (N=3630) and in each of these two populations are summarized below.

Just those	receiving	Bda a	nd treated	in 2015	(N=1210)

Past	N	Res H	Res R	Res Cm	Res	Res SLI	Res Ofx	Res Lfx/Mfx
Treatment	11	(%)	(%)	(%)	Am/Km (%)	(%)	(%)	(%)
FLD	487	92.2	100	24.4	31.8	33.9	32.2	7.4
SLD	251	98.0	100	46.6	57.8	61.8	68.1	20.7
Unknown	472	89.6	100	25.8	32.6	35.2	34.5	4.9

or None			
---------	--	--	--

Number with XDR = 340 (28%)

#### Just the randomly selected Controls (No BDQ, treated in 2015; N=2420)

Past	N	Res H	Res R	Res Cm	Res	Res SLI	Res Ofx	Res Lfx/Mfx
Treatment	11	(%)	(%)	(%)	Am/Km (%)	(%)	(%)	(%)
FLD	998	69.2	100	4.6	4.7	6.2	4.6	0.8
SLD	175	83.4	100	18.8	25.7	28	28	0.6
Unknown or None	1247	71.85	100	4.57	5.37	6.82	4.65	0.7

Number with XDR = 96 (4%)

For BDQ specifically the SA data will be used in two ways, as this data set represents the largest experience with BDQ anywhere in the world to date.

The primary BDQ analysis will include the 1210 treated with BDQ in 2015. These will be combined with all those treated with BDQ in all other cohorts included in the 2018 IPD.

<u>Secondary BDQ analysis - Case control approach</u>: will involve SA data only and use a case control approach. We will select the 1344 SA patients who received BDQ and started MDR treatment in 2014 and 2015 (note: 32 patients received bedaquiline between 2010 and 2013; they will be excluded). Each BDQ treated patient will be matched exactly (except age where they will be matched to 'nearest neighbor') to two controls: Historical controls treated in 2010-11 (HC) and concurrent controls treated in 2014-2015 (CC), on the basis of;

Province: All 9 listed Age: Continuous

Resistant to any SLI = Yes, No Resistant to any FQ = Yes, No

AFB Smear = Positive, or, Negative/Unknown

HIV = Positive, or Negative/Unknown

Previous Treatment= New/Prior FLD, or Prior SLD

For the 1344 patients with MDR-TB receiving BDQ in 2014-15, there were 9315 patients in these same years who did not receive BDQ to match to. After running the matching algorithm 1209 BDQ patients (90%) were successfully matched to other patients in 2014-2015. This same process was completed for the 6569 patients treated in 2010-2011 who did not receive BDQ. After running the matching algorithm 1071 BDQ patients (80%) were successfully matched. Upon combining these two cohorts together, there were 1050 BDQ patients (78%) with two controls (i.e. one from the 2014-2015 cohort and one from the 2010-2011 cohort), 21 BDQ patients (2%) with only one control in the 2010-2011 cohort, and 159 BDQ patients (12%) with only one control in the 2014-2015. Taken together, of the 1344 BDQ patients that could be matched to a control, 1230 (92%) were successfully matched to at least one control.

Additional Sensitivity BDQ analysis: Exclude SA data – use all other data of patients who received BDQ in the IPD

#### 1.2.2 Data received from Otsuka:

We have received three data sets from Otsuka: (i) Results for 9 patients treated under the Compassionate Release program; ii) Trials 204/208 and 116; 3) Trial 213.

(i) The 9 patients were highly selected; all of them had been treated for many years and in all of them, delamanid was the single drug added. We felt that these individuals represented a very unusual patient population; these were not included.

#### (ii) Trial 204/208/116

Trial 204 involved 481 patients who were randomized to two months of delamanid or placebo, in addition to an optimized background regimen. Of these 481 patients, we received a dataset for 421 (participants in Trial 116) for whom there were known end of treatment outcomes. 60 individuals did not participate in Study 116 – for these patients end of treatment outcomes are unknown, although we do know their vital status ie. whether they died or survived up to 30 months post randomization. Of the 421 for whom we have more complete data, 230 patients entered study 208 in which they received an additional 6 months of delamanid. The patients who participated in study 208, when compared to patients who participated in study 204 and 116 but did not participate in 208, had much lower death, and default/drop-out rate but had significantly higher failure rate. Note these patients were self-selected and/or selected by their providers to receive the additional 6 months of delamanid, after a gap of several months between completing the first 2 months of study 204 (the initial 2 months of delamanid versus placebo), and then starting the added 6 months of delamanid. This could have resulted in very substantial selection bias of patients into study 208 including a "survival bias" in that patients had to survive plus be adherent with treatment to be eligible for study 208. I believe this selection bias accounts for the low death and default rates as well as high failure rates in this sub study.

In addition, to this potential for selection bias, and unknown outcomes for almost 12% of the initial population, we have very little information on the MDR treatment received, in addition to Delamanid. We have only information on the drugs received at the time when patients were receiving delamanid or placebo in study 204 (in the first 2 months of treatment), or when receiving delamanid as part of study 208. This means that for more than half the patients, we have only information on drugs received in the first 2 months and then for another approximately 40% of patients we have information on drugs received for another 6 months, and a separate time. A final problem with this study data is that the manufacturer (Otsuka) will not release the information regarding whether patients were initially randomized to placebo or delamanid but will only provide to us whether patients received 0-2 months of delamanid or 6-8 months of delamanid. Hence, we cannot analyze individuals who received no delamanid whatsoever and compare them to patients who received any delamanid. For all these reasons this data set has been excluded from all analyses.

#### (iii) Trial 213:

For PICO2 of DLM we will analyze the data from the sponsor – and produce estimates of the effect of DLM on end of treatment outcomes – fail/relapse vs success and death vs success, as well as 2 and 6 month culture conversion (binary outcome), and acquired resistance to SLI or FQ. In secondary analyses, we will pool the data from trial 213 with all other patients who received delamanid in other datasets (notably Latvia, France and END TB). This will consider the trial 213 patients who received delamanid as another cohort, to examine end of treatment outcomes, specified above, with delamanid. We will show the results of Trial 213 as a distinct trial in Grade tables for this PICO (ie as a separate analysis for this one RCT).

#### 1.2.4 Assembling the final 2018 IPD data set:

The final 2018 IPD data set will include data from the 2016 IPD, plus the new data sets as above, including the sample of patients records from 2015 from South Africa. We will not include the Trial 213 data from Otsuka in all

the 2018 IPD analyses as we have permission to use this data only for analysis of PICO2 for Delamanid. Hence these records will be excluded from the general IPD used for all other PICO questions and will be excluded as a cohort during other drug-specific analyses within PICO2.

We will also exclude from the 2018 IPD seven data sets that were included in the 2016 IPD. In these 7 data sets the majority of patient records were missing information of DST to FQ and SLI. Six were programmatic data sets with 3156 patients (from Botswana, Brazil, Dominican Republic, Mongolia, Nepal, and Taiwan) and were missing DST to these two key classes of drugs in 61% to 100% of patients. All used standardized regimens. In addition, a data set from a referral hospital in Iran with 211 patients was missing DST to FQ and SLI in 67% of patients. Since these two variables are key determinants of prognosis and response to therapy, and used for PS matching – we will exclude all 7 data sets from the 2018 primary IPD analyses. But these 7 data sets will be included in sensitivity analyses for PICO2. Given the missing DST information for key drugs these 7 data sets can NOT be used for any analyses of PICO3 (N drugs), but can be used for sensitivity analyses of PICO4-6 (Duration).

After merging all records we have a **2018IPD** data set of **13,129 patients**.

From this we will exclude:

**11 patients** who received only 1 or 2 total drugs. We believe this is implausible, and so may represent errors in information.

**14 patients** received <6.0 months total duration of MDR treatment, yet were listed as having treatment success. We believe this is implausible as well, and likely represents incomplete information.

After these exclusions – we have a final 2018IPD population of 13,104 patients

#### 1.2.5 Selecting appropriate comparators for each PICO question:

The PICO questions generally refer to a comparator regimen that is concordant with WHO recommendations (minimum 18 months total, 8 months injectable and 4 likely effective drugs plus PZA or 5 likely effective drugs). Therefore, for each PICO question the comparator group will be a subset of the total **2018IPD** population. The specific parameter definitions selected for the comparator populations for each PICO question, and the resultant size of the total population eligible for the PICO analysis are summarized below.

#### 1. Minimum Total Duration:

Patients with success (comparator group for fail/relapse and for death) must have received at least 17.5 months. A substantial number of patients with success had therapy for 17.5 - 17.99 months, which we assume may represent such things as the last visit date fell before actual completion of therapy, and we believe these patients should be considered to have received at least 18 months of drugs. (95 patients with success had duration of 17.0 to 17.49 months – these are excluded). In the 2018 IPD we have 5 centres where >25% of patients with outcome of treatment success received less than 17.0 months total, although the mean treatment duration in those achieving cure/complete/success is less than 17 months in only 3 of these studies (LatviaDlm2018 [16.9 months], van der Werf, and Bang [both 16.7 months]). In these centres, we will apply the same rule and exclude only the individual patients with success who had total duration <17.5 months for the primary analysis for PICO 2-3.

For PICO3 - This would reduce our analyzed total population - from 13,104 to 12,553.

#### 2. Initial intensive phase duration:

This will NOT be used for selection of comparator (many patients received no injectable and this is an increasing phenomenon. Some studies did not specify duration of initial phase – hence this criterion will not be applied

### 3. Minimum number of possibly effective drugs:

For this analysis – only, a drug will be counted as effective if given EVER during treatment and they are confirmed susceptible on DST, or their susceptibility is unknown (no imputation, unknown=susceptible). For the calculation of number of possibly effective drugs, if an individual received more than one injectable, this is only counted as ONE; same for fluoroquinolones; same if they received Cs and Trd or Eto and Pto, or Ipm and Mpm. We also do not count H, R, Rfb, macrolides, AmxClv (by itself) or High dose INH as effective. We will keep patients as comparators (success) if they received at least 4 effective drugs plus PZA (regardless of DST result to PZA), or at least 5 effective drugs (if no PZA)

**For PICO 4-7** – this would reduce our analyzed total population by 2,475 from 13,104 to **10,629.** Note that this will exclude 731 patients with XDR (meaning 38% of the total of 1917 with XDR in the full IPD) from these analyses.

For PICO 2 - we must limit the comparator group to those with success who received the required minimum duration and minimum number of drugs (ie combine the rules). We reduce the study population by 2874 individuals, bringing the total population analyzed to 10,230.

### 3. Study measures and definitions

Main exposure variables

<u>Drug(s) Used</u> (and definition of effectiveness based on DST) – we consider a drug used if it was administered for at least one month (30 days) during the MDR-TB treatment episode. We will describe the mean and SD (or: median and IQR), duration of each drug used (including number of patients and number of studies in whom this is known), in the tables of results. The minimum duration for a drug to be considered 'used' will be further assessed in sensitivity analysis (Section 6). Note that where available DST information will be used for these analyses. Analysis of each drug will be stratified by DST: Drug x use among patients with isolates susceptible or resistant to Drug X. For most drugs we will exclude patients with DST missing for that drug from analyses of efficacy of that drug.

For certain drugs we will assume the isolates are susceptible if the DST are missing, if the prevalence of resistance to the drug is <10% in the overall IPD and/or published studies. These drugs include: Carbapenems, BDQ, DLM, LZD, CFZ. In updated analyses using the new 2018 IPD, 50% of patients had available DST results for Cycloserine/Terizidone (Cs/Trd); of these only 8% of isolates demonstrated resistance, so if DST to Cs/Trd is missing these isolates will be assumed susceptible.

DST to Amikacin and Kanamycin will be considered equivalent, if only one of these two is available. Similarly, DST to Cipro or Ofx will be considered equivalent, and to Mfx or Lfx will be considered equivalent. If DST to Cfx/Ofx is available but not DST to Lfx/Mfx – and Lfx or Mfx is used, then effectiveness of the drugs will be based on the DST to Cfx/Ofx. If Lfx or Mfx are used and DST to Lfx or Mfx is available – then effectiveness will be based on that more specific DST.

MDR regimens have multiple drugs, but for certain drug classes, such as the injectables or the fluoroquinolones (FQ), only one drug within that class should be used at any one time. Some patients have received more than one injectable or more than one FQ. For most of these, we know dates that individual drugs were started and stopped, so can determine which was used predominantly. However, if one patient received two different injectables or FQ – each for more than 1 month – then they will be excluded from the PICO2 analysis of both drugs. If use of one

drug is less than 1 month, and another is more, then we will disregard the drug used for less than 1 month. If patients received Streptomycin, then a second line injectable *AND* their isolate was resistant to Streptomycin plus sensitive to the SLI, then the Sm will be disregarded, and only the SLI analyzed. Similarly, if they switched from Ofloxacin to Levo or Moxifloxacin *AND* their isolate was resistant to Ofx but sensitive to the later gen FQ, then their results will be analyzed with Levo or Moxi (not Ofx).

#### Definitions of specific key variables

<u>Number of Effective Drugs Used</u> – drugs will be counted as effective if DST was done and showed susceptibility, or not done, but prevalence of resistance to that drug was <10% in the same population (from same country or study site if within one country, or overall at all sites if local data not available). Based on WHO criteria, the reference group will be 4 likely effective drugs  $\pm$  PZA. Other subgroups, adding or taking likely effective drugs, will be evaluated. These are subject to change based on sample size in each group. See below for sensitivity analysis (Section 6) in relation to drugs stopped due to AE and thus, likely double counted as an effective drug.

<u>Intensive Phase Duration</u> – this will be defined as the duration that an injectable is used. If no injectable is used, then that patient will be excluded from this analysis, *even if the investigators defined an initial intensive* phase (based on number of drugs for example). In addition to standard analysis using our pre-specified control group, to assess the optimal duration of the intensive phase of treatment, generalized additive mixed effects models using a spline (cubic regression) with a minimum of 5 splines will be explored. This will (1) allow the relationship between the outcome of interest and duration to be non-linear and (2) allow the data to define the optimal duration to maximize (or minimize) the odds of an outcome of interest.

<u>Total Duration of Treatment</u> – this is the time between treatment start and treatment end, as reported in each center for each patient. Treatment start is defined by the authors/investigators, or, when unclear by the date of start of at least two second line drugs. In addition to standard analysis using our pre-specified control group, to assess the optimal duration of treatment, generalized additive mixed effects models using a spline (cubic regression) with a minimum of 5 splines will be explored. This will (1) allow the relationship between the outcome of interest and duration to be non-linear and (2) allow the data to define the optimal duration to maximize (or minimize) the odds of an outcome of interest. Total treatment less than 6 months, if coded as complete or cured, will be excluded.

<u>Use of Sputum Culture</u> – this will be considered present if a patient has  $\geq 1$  sputum culture result reported **post-baseline**.

### Outcome variables

Laserson et al. [2] or the WHO Definitions and Reporting Framework [3] define end of treatment outcome variables for Tuberculosis 2013.

Other outcomes (acquired drug resistance and culture conversion) are defined below:

<u>Cure</u> – Laserson: an MDR-TB patient who has completed treatment according to country protocol and has been consistently culture-negative (with at least five results) for the final 12 months of treatment. WHO 2013: Treatment completed as recommended by the national policy without evidence of failure AND three or more consecutive cultures taken at least 30 days apart are negative after the intensive phase.

<u>Treatment completed</u> – an MDR-TB patient who has completed treatment according to country protocol but does not meet the definition for cure or treatment failure due to lack of bacteriologic results.

<u>Success</u> – Cure and Treatment completed will be considered together as treatment success.

<u>Failure</u> – Laserson: treatment will be considered to have failed if two or more of the five cultures recorded in the final 12 months are positive, or if any one of the final three cultures is positive, or if a clinical decision has been made to terminate treatment early due to poor response or adverse events. WHO 2013: Treatment

terminated or need for permanent regimen change of at least two anti-TB drugs because of: lack of conversion by the end of the intensive phase, or bacteriological reversion in the continuation phase after conversion to negative, or evidence of additional acquired resistance to fluoroquinolones or second-line injectable drugs, or adverse drug reactions.

<u>Relapse/Recurrence</u> – an MDR-TB patient who was declared cured or treatment completed at the end of the MDR-TB treatment, and was diagnosed with a recurrent episode of MDR-TB after the previous episode (in the primary analyses, failure and relapse will be considered together). If information is available to distinguish reinfection from relapse (i.e. using molecular markers), then reinfections will be excluded, and only true relapses counted. However, in the great majority of datasets this information is not available – in these, all recurrences will be counted.

<u>Death</u> – an MDR-TB patient who dies for any reason during MDR-TB treatment.

<u>Acquired Drug Resistance</u> – Will be defined as new resistance on DST performed after at least 3 months of MDR therapy, and compared to baseline or pre-MDR treatment.

Culture Conversion — where culture is available and a patient begins treatment culture positive, a patient will be considered to have converted when two consecutive sputum cultures at least 28 days apart are negative; the culture conversion date will be date of sample collection of the earlier of the two negative cultures. If a patient begins treatment as culture negative and has ≤1 positive sputum culture in the first three months, they will be excluded from any analysis of culture conversion. In cases where a patient's final culture is negative, but is not preceded by another negative culture, the culture conversion date is considered missing. In other cases, the patient will be considered to not have converted. A sensitivity analysis for this outcome is planned (Section 6). Lost to follow-up — includes dropout, patient decision to stop therapy, transferred out without known outcome, defaulted, and not accounted for. Loss to follow-up, in prior analyses, was strongly associated with centre. This may reflect different resource levels available to support patients throughout MDR treatment. As such there may be strong confounding of centre level support for adherence and availability of certain drugs (e.g. BDQ, LZD, certain injectables). Prior to beginning the analysis, we will analyze the main determinants of LTFU as described in section 5 below. This analysis will inform decisions regarding inclusion of this outcome in analyses for PICO 2 and 3. LTFU cannot be analyzed with regard to duration, as the outcome (LTFU) determines the duration (like death).

#### Adverse events

 Adverse events are being assessed in a separate analysis of the 2016 IPD database, and will not be analyzed using data provided in these new patient populations.

#### Covariates

- Demographics: Age, Sex, HIV-status. (We will examine association of outcomes and treatment with other
  comorbidities or habits such as diabetes, or smoking if adequate numbers of patients to permit these
  added analyses)
- Disease Severity Indicators: Acid- fast bacilli (AFB) smear positive, Cavitation on chest X-ray
  - This will be defined as "extent" of disease, first based on AFB smear. If that status is unknown, then based on cavitation on chest X-ray; if unknown, then bilateral disease.
- Drug Susceptibility: Resistance to a FQ, resistance to a second-line injectable (SLI)
  - These will be excluded when evaluating specific FQ or SLI in PICO2 (i.e. we will not match on FQ-R when evaluating a drug like Mfx).
- Previous Treatment: First-line drug (FLD) treatment, second-line drug (SLD) treatment
- Treatment Regimen: Number of effective drugs in the initial (intensive) phase
- World Bank Income bracket for country: Three categories (Low&Low-middle, Upper middle, and High)

### 4. Missing data imputation

For analyses requiring the above covariates, missing values of covariates (but not the primary independent variables of interest in any such analyses) will need to be imputed. We will use multiple imputation (*mice* in R), as the variability around parameters is better considered, and the resultant confidence intervals are appropriately broad (mean value imputation underestimates the uncertainty and hence overestimates the precision).

Missing DST information: In the new 2018 IPD, 95.2% of patients have information on drug sensitivity testing for at least one fluoroquinolone (FQ) and 95.4% have DST information on at least one second line injectable. 4.3% had no DST information for any FQ nor any SLI, and 0.7% had information on one or the other class of drugs but not both; 95% of patients had DST information for both major classes of drugs. For the other drugs DST information was more often missing. Note that if DST was missing for clofazimine, linezolid, bedaquiline, delamanid, or carbapenems, we assume the isolate was sensitive because in the entire dataset, less than 5% of isolates tested were resistant to any of these drugs.

In previous analyses, we have used the 10% rule that if, among all those individuals with DST results, less than 10% are resistant, that we assume the isolate is sensitive, if DST not available. Applying this rule will mean that isolates would be considered susceptible to cycloserine or terizidone if DST missing.

Imputation for missing DST information for FQ, SLI, Sm, PZA, EMB, Eto-Pto, and PAS, will be performed using: Age; Sex; HIV; Extent of disease (based on AFB smear, if missing, then on CXR cavitation, then on bilateral, then if all missing, then imputed); Previous Drug exposure (None, FLD, SLD), Year, and DST information (same drugs)

Imputation for missing Age, Sex, Prior drug exposure, HIV will utilize: these same variables, plus DST information, outcomes and country Income Level (three level, as described in covariates)

Imputed values will be generated 20 times. Each of these data sets will be analyzed for each PICO and then combined to arrive at an overall result (aOR or aRD) according to Rubin's rules.

Note that imputed variables will only be used for propensity score matched analysis, and not for any primary analysis. For example: For PZA and outcomes we will stratify by known PZA resistance; patients will be included in these analyses only if the DST to PZA is known.

Imputing extent of disease: Chest x-ray information is available for only about 50%. AFB smear information is available for more than 80%. Hence, we will use AFB smear information as the primary determinant of extent of disease. If AFB smear positive, the individual will be classified as having extensive disease and if AFB smear negative as not having extensive disease. If AFB smear is unknown, then chest x-ray cavitation will be an indicator of extensive disease. If information on cavitation is also missing, then bilateral disease on CXR will be used. If there is neither AFB smear nor chest x-ray information, then we will impute the extent of disease using the above methods.

#### 5. Loss to Follow-up Analysis

Inclusion of individuals who have been lost to follow-up (LTFU) can be problematic as they may be systematically different from others, but even more importantly, this outcome may be strongly associated with site-related characteristics. To examine this, prior to analysis of this outcome for PICO questions 2 and 3, we will assess whether, *after* accounting for Site, and individual patient characteristics, there is any effect of drug on LTFU. We will do this as follows: create a generalized linear mixed model with no explanatory variables, a random intercept by site, and the outcome of LTFU. We will then adjust for patient-characteristics associated with LTFU (e.g. age, sex, presence of co-morbidities, cavitation) and see how this affects the variance associated with

the random intercept. In theory, we expect this to significantly account for much of the variability, but not all of it. As a third step, after accounting for site and individual characteristics, treatment characteristics that are largely common to all sites, and therefore not correlated with the sites themselves, (e.g. treatment with pyrazinamide, ethionamide) will be examined for correlation with LTFU. If the inclusion of these patient- and treatmentcharacteristics completely explain the variance seen in the site (random intercept), we can conclude that LTFU is not due to unknown site-related factors and therefore can be included in analysis. If variance remains, then we will exclude LTFU from primary analyses.

#### 6. Effectiveness analyses

Effectiveness analyses have been conducted and reported before, using the 2010 IPD, 2016 IPD and short regimen IPD. Newly acquired data will be merged with the 2016 IPD and meta-analyzed. Evaluation will be done in the 'group of interest' compared to the 'control group' as defined in Section 3 where exposure variables are outlined. An example: for analysis of specific drugs, the group of interest will be those who received the drug and the control group will be those who did not receive the drug.

The analyses will be repeated for each patient population of interest, as defined in **Section 2**. The outcomes evaluated will be the adjusted odds ratio (aOR) and the adjusted risk difference (aRD). These will be calculated in terms of unfavorable outcomes (e.g. the *lower* the odds, the better, and *negative* risk differences reflect better outcomes). We are proposing two outcomes, as outlined below, separated for analytical purposes.

Odds of Failure and Relapse:  $\frac{Failure + Relapse}{Cure + Complete}$  Risk of Failure and Relapse:  $\frac{Failure + Relapse}{Failure + Relapse + Cure + Complete}$ 

Odds of Death:  $\frac{\textit{Death}}{\textit{Cure} + \textit{Complete}}$ 

Risk of Death:  $\frac{Death}{Death + Cure + Complete}$ 

The odds ratio and risk differences will be calculated for the group of interest compared to the control group (as previously defined).

We could consider a composite outcome of death plus fail/relapse as outlined below:

Odds of Unfavorable Outcome:  $\frac{Death + Fail + Relapse}{Cure + Complete}$ 

However, we do not favor this approach for two major reasons. First, in a composite outcome the outcomes combined should be reasonably equivalent. Death is not equivalent to fail/relapse (ask any patient). Secondly, drugs may act in different ways and so produce different effects. A drug that acts quickly to reduce disease burden may reduce mortality without necessarily affecting the longer-term failure/relapse. Or a drug may have important adverse effects that increase mortality (e.g. Bedaquiline in early trials) – which would be important to detect.

Pending analysis of loss to follow-up (see Section 5) the outcome of loss to follow-up might be included as part of an additional composite outcome:

Odds of Unfavorable Outcome: 
$$\frac{\textit{LTFU} + \textit{Death} + \textit{Fail} + \textit{Relapse}}{\textit{Cure} + \textit{Complete}}$$

Or, loss to follow-up might be analyzed separately:

Odds of LTFU: 
$$\frac{LTFU}{Cure + Complete + Fail + Relapse}$$

Or, time to LTFU could be analyzed in a form of survival analysis. In this death and failure would be considered censored when they occur (meaning a patient would be 'at risk' of LTFU up until they failed or died.)

# All effect estimates will be adjusted for covariates using exact matching for three variables, and propensity score matching for others.

Given the very important impact of FQ and SLI resistance on outcomes, and response to therapy, we will match **exactly** on these two parameters (albeit *any* FQ, and *any* SLI resistance). Since we will exclude all data sets in which these two parameters are missing in more than 25% of patients (see again Sections 1.1 and 1.2 above), this will not result in an appreciable reduction in number of patients. We will also match exactly on country income level (from World Bank – 3 categories: Low&lower-middle, upper-middle, and high-income)

A propensity score is the probability of allocation to treatment given the measured covariates [5]. Propensity score for each individual will be calculated using logistic regression (*matchit* in R), based on the covariates described above. We will use propensity score matching on HIV, age, sex, prior TB drug exposure (none, first line, second line), extent of disease, and the number of effective drugs used in the initial intensive phase. Note that the number of effective drugs will now be based on the imputed drug sensitivity testing for many patients. This will increase the estimated number of effective drugs used, because previously, if DST information was not available, then patients receiving many of the drugs were considered not to have received an effective drug. Patients will be matched from the control group (randomly from those falling within the specified caliper distance) to the group of interest according to their propensity scores [6].

The matching method used will be one-to-one matching with replacement. This algorithm will use 'random' nearest neighbor matching to identify pairs, whose propensity scores differs by no more than a specified amount (the caliper distance)—this results in a random individual falling within the caliper distance to be selected. The adequacy of matching will be evaluated based on the comparison of covariates balance between matched groups. This process will be repeated 20 times using different "seeds" for the random number generator utilized in matching—this will likely result in a different individual selected in each imputation set, better capturing variability. The results of these 20 replications will be pooled according to Rubin's rules to arrive at an 'average' adjusted odds ratio (or risk difference) and a corresponding 95% confidence interval.

In secondary analyses, for covariates that we judge as very important (such as XDR) we can restrict to only patients with that characteristic (for example – effect of BDQ within XDR patients only).

LZD, BDQ, CFZ and Carbapenems are used in a limited number of centres, and within those centres to selected patients only. In planned sensitivity analyses for PICO2 for these 4drugs we will restrict the comparators to patients treated at other centres where the drug of interest is not used at all or used in <2% of patients.

#### All efficacy analyses will be completed using generalized linear mixed models.

A generalized linear mixed model (*lme4* in R) combines fixed effects and random effects in the modeling procedure. As covariates will be controlled for through matching, they will not need to be explicitly modeled. Instead analyses will be adjusted by using random intercepts. Random intercepts will be defined for matched pairs and data sets (i.e. sites)—random slopes will not be used as these cause significant convergence issues, based on previous experience. If models fail to converge with specification of these random effects, then preference will be given to only defining a random intercept for matched pairs [6]. Odds ratios will be calculated via the binomial function with a logit link, while risk differences will be calculated via the binomial function with an identity link (*blm* in R).

Due to the large numbers of factors on which we adjust or match, we will not perform PS matched analysis in sub-groups of less than 100 patients (in total).

## SENSITIVITY ANALYSES (in addition to those mentioned earlier such as for SA data)

# 1. Association of duration of treatment with relapse (failures excluded) in studies for which this is available (currently 16 studies from the 2016 IPD with only 2163 patients). (PICO 4-6)

Explanation: Treatment duration was not randomized. Any analysis adjusting for treatment duration thus suffers from imputation of **intended** treatment duration in failures as failures may be diagnosed before the usual, or planned end of treatment. Thus, a potential bias introduced is that shorter treatment may be associated with failure. Analysis of relapse (an outcome that by definition must occur *after* successful completion of treatment) in studies that routinely followed **all patients** after cure or completion for relapse, will eliminate this bias. (We will assess this – studies that followed only a small proportion of patients might have biased results as they may have selectively followed only those considered at higher relapse risk).

# 2. Association of duration of treatment post culture conversion with end of treatment outcomes, and timing of culture conversion (i.e. <2 months, $\ge2$ and <4 months, $\ge4$ and <6 months). (PICO 6)

<u>Explanation</u>: Timing of culture conversion can impact subsequent clinical care (e.g. extending treatment, or adding drugs because of delayed conversion). This may lead to bias in our analysis. By stratifying based on timing of culture conversion we can assess if the same results would be achieved across strata. If they are, we can conclude that there is less risk of bias in the analysis.

# 3. Restriction to patients with previous FLD exposure or no previous drug exposure (patients with previous SLD exposure excluded). (PICO 2-6)

<u>Explanation</u>: In some datasets, individuals with extensive prior treatment are candidates for use of specific drugs (e.g. BDQ or DLM), which may reduce estimated benefits of these drugs. As well, this makes this group more similar to those who received the short regimen. This analysis will exclude these patients to examine the impact on treatment outcomes – specifically improved favorable outcomes.

4. Patients receiving other new drugs – Linezolid, clofazimine, carbapenems and bedaquiline - are excluded. (PICO 2)

<u>Explanation</u>: Use of the new drugs was highly correlated, making it difficult to distinguish the effect of one d]new drug from the others.

5. Outcomes in patients with number of effective drugs used 3 + PZA (PICO 2)

<u>Explanation</u>: About 1,000 patients received treatment with fewer effective drugs than the minimum recommended by WHO. This large group was added back into PICO2 analyses.

6. Changing the definition of drug used from minimum 30 days (current base case analysis) to a minimum of 90 days, or minimum of 180 days. Restricted to datasets with information on duration of each specific drug (PICO 2, and PICO 3)

<u>Explanation</u>: Under the current framework for analysis, all drugs are counted as used if they were received for at least 30 days. But one month seems too short, so when duration of specific drug use is available, we will vary the minimum duration of use for a drug to be counted as 'used' to explore the potential impact of duration of each drug used on outcomes. Note: prior to analysis, we will describe the frequency of number of effective drugs used based on minimum duration of use.

7. Compare overall outcomes, and estimates of effects on outcomes of a few drugs of interest (e.g. Kanamycin, PZA, FQ and LZD) – in centres/studies using the Laserson vs WHO 2013 treatment outcome definitions. (PICO 2)

*Explanation*: This proved impossible – all studies used WHO2008. The only exception was Trial 213 – this was done for this PICO2 analysis.

- **8. Number of highly effective drugs:** we will count only Later generation FQ (if DST sensitive), LZD, BDQ and Carbapenems.
- **9. Restrict to centres using LZD, BDQ, CFZ and Carbapenems for more than 2% of patients:** These drugs were used in a limited number of centres, and within those centres to selected patients only. In planned sensitivity analyses for PICO2 for these 4 drugs we will restrict the comparators to patients treated at other centres where the drug of interest is not used at all, or used in <2% of patients.
- **10. XDR Sub-group analyses:** For all PICO questions we plan to repeat within the sub-group with XDR only.

#### 7. Assessing heterogeneity

Compared to traditional aggregate data meta-analysis, IPD meta-analysis, by definition, will have smaller between-study heterogeneity because of the ability to adjust for individual patient characteristics (covariates). Heterogeneity between the studies will be assessed by I² using a simulation-based method (developed and published by Dr. Benedetti [7]) and tau-squared (Tau²), using both unadjusted and adjusted effect estimates from each study (SAS v9.4).

**<u>8. Analysis software</u>**: SAS software v9.4 and R

#### 9. Quality assessment of studies

<u>This has been completed for all data sets previously included in our database</u>; we will follow the same methods for quality assessment of all new data sets provided. See Appendix for quality assessment completed of the studies in the 2016 IPD.

There are no published criteria for evaluating quality of IPD studies. We developed a checklist of 7 indicators, adapting from Risk of Bias In non-randomized Studies of Interventions (ROBINS-I) [8]. A study that achieves high quality in at least 6 out of the 7 indicators will be considered as high-quality study. A study that achieves high quality in 4 or 5 indicators will be considered moderate quality.

- Note: Two randomized trials were included in the 2016 IPD database, but all arms of these trials were
  analyzed as cohorts since the randomization in these trials applied only to one comparison. These were
  treated as cohort studies for quality assessment this same process will be followed for new trials
  included in this update. Please see the initial PICO question clarifications for how drugs included in these
  trials will be handled for analysis.
- (1-2) <u>Bias in selection of participants into the study</u>: if participants were selected based on certain characteristics before treatment (e.g., more likely adherent, or more extensive resistance) or after treatment start (e.g., excluding patients that stopped certain drug early), selection bias could be introduced. Selection of participants will be assessed based on the sampling method and participation rate in each study.
- What is the sampling method in the study: census (all patients), random sampling, or convenience sampling?
- If census sampling was the participation rate in the study >80%?
- (3) <u>Bias in measurement of outcomes:</u> pooling studies with different classifications of treatment outcomes could lead to information bias (misclassification) in the analysis. Thus following a standard method of reporting treatment outcome is a quality indicator. However one study has noted significant differences in outcome classification for the same patients using the two criteria [9]. Hence, a planned data quality analysis is to evaluate the impact of treatment outcome definition (i.e. Laserson vs. WHO 2013) on outcomes, and estimates of effect. To complete this, the studies will be stratified based on the outcome used in that study, as reported in the manuscript or by the investigators. We will compare overall outcomes of studies in the two groups (using aggregate pooling techniques), and also assess the association of four drugs (kanamycin, PZA, later generation FQ, and LZD) with outcomes to assess the potential influence of these two different definitions on our findings.
- (4-7) <u>Information bias of confounders</u>: (lack of information on key confounding factors): Age, HIV, previous TB treatment, and resistance to second-line drugs all have major impacts on treatment response, and so may be considered major confounders in the analysis of drug efficacy and drug adverse event. Therefore, adequate reporting of these confounders in each study is important for calculating unbiased effect estimates.
- Was age reported in at least 90% of the participants?
- Was HIV status reported in at least 80% of the participants? (If HIV prevalence is known to be less than 10% in TB cases in the study, or less than 1% in the general population in the country (i.e., low HIV prevalence), this item will be considered as high quality even if HIV status for individuals is not reported)
- Was previous TB treatment history reported in at least 80% of the participants?
- Were drug susceptibility testing results for at least one FQ and at least one second-line injectable reported in at least 80% of the participants?

#### 10. Revisions to SAP

This represents the SAP planned as of June 29, 2018. There may be revisions and additional analyses to manage remaining sources of bias, as applicable and necessary.

#### References

- 1. Ahuja SD *et al*. Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: an individual patient data meta-analysis of 9,153 patients. *PLoS Med* 2012. **9**(8): p. e1001300.
- 2. Laserson KF *et al.* Speaking the same language: treatment outcome definitions for multidrug-resistant tuberculosis. *Int J Tuberc Lung Dis* 2005;9(6):640-5.
- 3. World Health Organization. *Definitions and reporting framework for tuberculosis 2013 revision (updated December 2014).* 2014.
- 4. Hernan MA *et al.* Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol* 2006;98(3):237-42.
- 5. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* 2011;46(3):399-424.
- Fox GJ et al. Propensity Score-Based Approaches to Confounding by Indication in Individual Patient Data Meta-Analysis: Non-Standardized Treatment for Multidrug Resistant Tuberculosis. PLoS One 2016;11(3):e0151724.
- 7. Chen B and Benedetti A. Quantifying heterogeneity in individual participant data meta-analysis with binary outcomes. *Syst Rev* 2017;6:243.
- 8. Sterne JA *et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
- 9. Bastard M *et al.* Revised definitions of multidrug-resistant tuberculosis treatment outcomes: closer to the reality? *Am J Respir Crit Care Med* 2015;191(3):355-8.

Author Country	N (%)
Ahmad Pakistan	180 (1.37%)
_ Ahuja_USA	67 (0.51%)
Anderson_UK	132 (1.01%)
Bang_Denmark	21 (0.16%)
Barkane_Latvia	32 (0.24%)
BarryCalif_USA	53 (0.4%)
BarryKorea_South Korea	38 (0.29%)
BDQtrial_Brazil	2 (0.02%)
BDQtrial_India	5 (0.04%)
BDQtrial_Latvia	6 (0.05%)
BDQtrial_Peru	26 (0.2%)
BDQtrial_Philippines	3 (0.02%)
BDQtrial_Russia	7 (0.05%)
BDQtrial_South Africa	79 (0.6%)
BDQtrial_Thailand	4 (0.03%)
Bonnet_Georgia	204 (1.56%)
Brode_Canada	22 (0.17%)
Brust_South Africa	1261 (9.62%)
Cegielski_Estonia	57 (0.43%)
Cegielski_Latvia	148 (1.13%)
Cegielski_Peru	207 (1.58%)
Cegielski_Philippines	449 (3.43%)
Cegielski_Russia	139 (1.06%)
Cegielski_South Africa	521 (3.98%)
Cegielski_South Korea	110 (0.84%)
Cegielski_Taiwan	50 (0.38%)
Cegielski_Thailand	53 (0.4%)
ChanDenver_USA	15 (0.11%)
Dheda_South Africa	104 (0.79%)
endTB_Armenia	21 (0.16%) 12 (0.09%)
endTB_Belarus endTB Georgia	118 (0.9%)
endTB_Georgia endTB Kenya	1 (0.01%)
endTB_kenya endTB_Lesotho	14 (0.11%)
Fox Australia	26 (0.2%)
Gegia Georgia	380 (2.9%)
Guglielmetti2016 France	44 (0.34%)
Guglielmetti2018 France	10 (0.08%)
Hughes South Africa	52 (0.4%)
Isaakidis India	130 (0.99%)
Janssen Asia	83 (0.63%)
Janssen Europe	52 (0.4%)
Janssen Peru	18 (0.14%)
Janssen South Africa	77 (0.59%)
- <del>-</del>	•

Jarlsberg_USA	23 (0.18%)
Kempker_Georgia	150 (1.14%)
Koenig_Haiti	162 (1.24%)
Koh South Korea	336 (2.56%)
Kvasnovsky_South Africa	351 (2.68%)
Lange Germany	146 (1.11%)
Laniado-Laborin Mexico	50 (0.38%)
Leung Hong Kong	199 (1.52%)
Marks USA	120 (0.92%)
Migliori Belarus	9 (0.07%)
Migliori Belgium	15 (0.11%)
Migliori Brazil	3 (0.02%)
Migliori_Ecuador	5 (0.04%)
Migliori_Greece	15 (0.11%)
Migliori_Italy	99 (0.76%)
Migliori_Netherlands	49 (0.37%)
Migliori_Peru	23 (0.18%)
Migliori_Slovakia	3 (0.02%)
Migliori_UK	4 (0.03%)
MiglioriBDQ_Argentina	1 (0.01%)
MiglioriBDQ_Australia	1 (0.01%)
MiglioriBDQ_Belgium	2 (0.02%)
MiglioriBDQ_Greece	2 (0.02%)
MiglioriBDQ_India	16 (0.12%)
MiglioriBDQ_Italy	8 (0.06%)
MiglioriBDQ_Netherlands	2 (0.02%)
MiglioriBDQ_Peru	2 (0.02%)
MiglioriBDQ_Portugal	1 (0.01%)
MiglioriBDQ Russia	62 (0.47%)
MiglioriBDQ_South Africa	30 (0.23%)
MiglioriBDQ Sweden	3 (0.02%)
Milanov Bulgaria	49 (0.37%)
Ndjeka2016 South Africa	103 (0.79%)
Ndjeka2018 South Africa	3626 (27.67%)
ODonnell South Africa	111 (0.85%)
Palmero2015 Argentina	9 (0.07%)
Podewils Philippines	562 (4.29%)
Riekstina2016 Latvia	150 (1.14%)
Riekstina2018 Latvia	40 (0.31%)
Rodrigues Brazil	100 (0.76%)
Seo South Korea	25 (0.19%)
=	
Seung_North Korea	353 (2.69%)
Shim_South Korea	73 (0.56%)
Singla_India	29 (0.22%)
Skrahina_Belarus	98 (0.75%)

Smith_Russia	199 (1.52%)
Udwadia_India	18 (0.14%)
van der Werf_Netherlands	93 (0.71%)
Vasilyeva_Russia	125 (0.95%)
Viiklepp_Estonia	293 (2.24%)
Yim_South Korea	123 (0.94%)

Overall 13,104 (100%)

Narrative summary of evidence on Perchlozone, Sutezolid and Interferon-gamma for the WHO treatment guidelines for multidrug- and rifampicin-resistant tuberculosis (MDR/RR-TB), 2018 update

Prepared by Dr Kerri Viney (WHO External Consultant) Final version (version 4.0), finalised on 30 July 2018

**1. Background:** The *WHO* treatment guidelines for multidrug- and rifampicin-resistant tuberculosis (MDR/RR-TB), 2018 update will be informed by 7 questions formulated as PICOs (i.e. Population, Intervention, Comparator, Outcome). Of these, PICO 2 is concerned with the use of individual agents as part of a longer treatment regimen for MDR/RR-TB and whether the addition of these individual agents is likely to improve a number of pre-defined outcomes. The formulation of PICO 2 and the individual agents being assessed is provided in Appendix one. To assess the evidence on the use of Sutezolid, Interferon-gamma and Perchlozone (three of the less well known individual agents listed in PICO 2), a literature review was carried out in May-June 2018. These agents will not otherwise be included in the individual patient data (IPD) meta-analysis of MDR-TB treatment outcomes which is the main knowledge base driving the 2018 guidelines update.

The **objective** of the literature review was to source the latest available published evidence on the use of Sutezolid, Interferon-gamma and Perchlozone among patients with MDR/RR-TB to inform PICO 2 of the WHO treatment guidelines for multidrug- and rifampicin-resistant tuberculosis (MDR/RR-TB), 2018 update.

The **research questions** for the literature review were as follows:

- 1. Should sutezolid vs. no sutezolid be used for patients with rifampicin resistant or multidrug resistant TB (MDR/RR-TB) on longer regimens conforming to WHO guidelines?
- 2. Should perchlozone vs. no perchlozone be used for patients with rifampicin resistant or multidrug resistant TB (MDR/RR-TB) on longer regimens conforming to WHO guidelines?
- 3. Should interferon-gamma vs. no interferon-gamma be used for patients with rifampicin resistant or multidrug resistant TB (MDR/RR-TB) on longer regimens conforming to WHO guidelines?

#### 2. Methods:

**2a. Study selection and initial appraisal:** A search strategy was developed by in collaboration with specialist medical librarians from the Karolinska Institutet. Then, between May to June 2018, the following databases were searched: PubMed, Medline Ovid, EMBASE, Web of Science and Cochrane Library Database of Systematic Reviews. The database search was complemented by searching the grey literature, searching reference lists of highly relevant papers, searching relevant websites (such as UNITAID, Stop TB Partnership, Treatment Action Group, etc.) and contacting international experts (particularly for Perchlozone where the majority of publications are published in the Russian language). Professor Andrei Maryandyshev from Northern State Medical University (Arkhangelsk, Russia) conducted a separate search of a Russian database (ринц) for papers on Perchlozone and he contacted Russian experts for additional papers.

Duplicate studies were removed and then selected studies were first screened using the title and abstract, assessed for eligibility based on the criteria outlined in PICO 2. A review of the full text papers was then conducted. Papers retrieved in the Russian and Chinese languages were reviewed by native Russian (Professor Maryandyshev) and Chinese (Dr X Yinyin at WHO) speakers. Data from

the selected full text papers were extracted into a summary worksheet. The results of the search strategy are summarised in three PRISMA flow diagrams, one for each individual agent (Appendix two).

**2b:** Appraisal by the Guideline Development Group: The methodology and preliminary results of the literature review were presented to the Guideline Development Group (GDG) on 30 May 2018 in one of the GDG's regular webinars. The presentation also included information on a separate literature review on the cost effectiveness of individual agents used in MDR/RR-TB regimens. There may be some slight discrepancies in the number of papers assessed in the presentation to the GDG and this summary report due to the fact that a small number of pre-ordered papers became available after the GDG webinar on 30 May 2018.

The GDG members were asked to comment on the findings of the literature review during the webinar. Based on the evidence presented the GDG did not deem it appropriate to construct a GRADE evidence table for Sutezolid or Interferon-gamma, based on the current limited evidence. The GDG did recommend that a GRADE evidence table be developed for Perchlozone for the outcome of culture conversion by 6-8 months. One study on Perchlozone provided evidence on adverse events, which was also included in the GRADE evidence table. The resultant GRADE evidence table is included in Appendix three.

**2c. Analysis:** For the four studies on Perchlozone with an outcome of culture conversion by 6-8 months, outcomes were pooled using random effects meta analysis of proportions (conducted in Stata, version 15 (StataCorp. 2017. *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC). A biostatistician from the Karolinska Institutet provided advice on the meta-analysis. The forest plot for the meta-analysis is included in Appendix four.

### 3. Results

**3a:** Literature review: A total of 326 articles were assessed (187 for Sutezolid, 94 for Interferon gamma and 45 for Perchlozone). The full text pdf was then retrieved for a total of 156 full text articles which were reviewed (60 for Sutezolid, 51 for Interferon-gamma and 45 for Perchlozone). Appendix two includes the PRISMA flow diagrams for the literature reviews for each agent.

**3b:** Current evidence on Sutezolid: Sutezolid (PNU-100480) is an oxazolidinone (a thiomorpholinyl analog of linezolid) which was first discovered in the 1990s.(1) It is currently licensed to Sequella Inc. (1) Based on the literature review, Sutezolid has been subject to Phase 1 and 2a trials only. A total of four papers describing Phase 1 trials (2-5) and one paper on the Phase 2a trial were retrieved (6). One of the Phase 1 trials was a modelling study based on the Phase 2a trial and both of these studies were concerned with drug susceptible TB only.(2, 6) All five studies were funded by Pfizer (to whom the drug was licensed prior to Sequella Inc.).

In the three Phase 1 trials described below, outcomes assessed included: whole blood bactericidal activity, pharmaco-kinetic parameters, safety and tolerability.(3-5)

In the first of these studies (the first study of Sutezolid in humans, published in 2010), the authors assessed the safety, tolerability, pharmacokinetics, and mycobactericidal activity of single ascending doses of Sutezolid against *M. tuberculosis*.(4) Nineteen people received escalating doses of Sutezolid, while 8 received Linezolid.(4) The authors reported that all doses of Sutezolid were safe

and well tolerated and that Sutezolid doses to 1000 mg were well absorbed and showed approximately proportional increases in exposures of parent and metabolites.(4) They concluded that single doses of Sutezolid to 1000 mg were well tolerated and exhibited anti-mycobacterial activity superior to 300 mg of Linezolid at steady state.(4) Minimum inhibitory concentrations and whole blood bactericidal activity of Sutezolid were also reported.(4) The authors concluded that additional studies are warranted to define the role of Sutezolid in the treatment of drug resistant TB.(4)

In a subsequent Phase 1 trial (described as the second study of Sutezolid in humans, published in 2011) the authors describe that the aim of this study was to determine a safe and efficacious dose of Sutezolid to advance in clinical trials and to determine the role of pyrazinamide. (3) Fifty subjects in five cohorts were randomly assigned to Sutezolid or placebo (4:1) in different dosing schedules for 28 days, to which pyrazinamide was added on days 27 and 28 for one cohort only. (3) A sixth cohort (of 8 people) were given Linezolid at 300 mg daily for 4 days. (3) The authors reported that "all doses were safe and well tolerated. There were no hematologic or other safety signals during 28 days of dosing at 600 mg twice daily." (3) Additionally, plasma concentrations and cumulative whole blood bactericidal activity were reported and the authors concluded that future studies are warranted and that drug development can be accelerated by the use of biomarkers. (3)

In the third Phase 1 trial, the authors used a single healthy volunteer as a blood donor and assessed whole blood bactericidal activity, hypothesizing that it may be "an emerging biomarker for TB treatment." (5) This study assessed the effects of several drugs, including Sutezolid, "TMC207, PA-824, SQ109, and pyrazinamide, singly and in various combinations, against intracellular *Mycobacterium tuberculosis*, using whole blood culture."(5) The study demonstrated that combinations of Sutezolid, TMC207, and SQ109 were fully additive.(5) For PA-824 they were less than additive or antagonistic.(5) The most active regimens, including Sutezolid, TMC207, and SQ109, were predicted to have cumulative activity comparable to standard TB therapy.(5)

The fourth Phase 1 trial on Sutezolid was a modelling study based on the Phase 2a clinical trial, both published in 2014. This study and the Phase 2a clinical trial were both tested in patients with drug susceptible TB. As the WHO guidelines are concerned with RR or MDR-TB they were deemed to be not relevant to the population described in PICO 2. There were no Phase 2b trials retrieved in the literature search and it appears that there have been no published papers on the use of Sutezolid since 2014. A search of ClinicalTrials.gov found reference to one trial that mentions Sutezolid as one drug of many in the intervention and comparator arms, however there is insufficient detail to be able to determine the role of Sutezolid in this trial, and it appears that Sutezolid is not a central component of the intervention.

**3c.** Current evidence on Interferon-gamma: Interferon-gamma is a soluble cytokine, essential for anti-mycobacterial host defences. There have been several small studies on the use of interferon-gamma as an adjunctive treatment for both drug susceptible and drug resistant TB.(7-11) The seminal study on the use of interferon-gamma among patients with MDR-TB was a study by Condos et al, published in the Lancet in 1997.(7) A total of 5 MDR-TB patients participated in this open label, non-randomised, prospective cohort study, where they were given aerosolised interferon-gamma

3

<sup>&</sup>lt;sup>1</sup> A brief description of this trial is available at: https://clinicaltrials.gov/ct2/show/NCT03237182?term=Sutezolid&cond=Tuberculosis&rank=5 (ClinicalTrials.Gov)

(500 μg thrice weekly for a month).(7) Safety, tolerability, signs and symptoms of TB and clinical and radiological parameters were the outcomes of interest.(7) This study showed that interferongamma was well tolerated by patients, sputum smears became negative in all patients, there was a non-significant increase in the time to positive culture (from 17 to 24 days reflecting the bacterial burden) and the size of cavitary lesions was reduced in all patients.(7) This study, however, did not include a control group and subsequent studies with small numbers of enrolled patients have not included these comparison groups either.(9, 10)

The literature review yielded a number of studies on interferon-gamma as an adjunct treatment for TB, although only two were included in the final summary of studies as many did not include a comparison group (and therefore we were not able to generate estimates of effect) or they had major flaws. No randomised controlled trials were found and it appears that the only randomised, placebo controlled, multi-centre trial of adjunctive interferon-gamma for patients with MDR-TB was the InterMune study which was halted prematurely due to deaths in the experimental arm. (12, 13).

Four studies were considered after the literature review was complete, two were published in the Chinese language and these papers were reviewed by a person who is a native Chinese speaker. Two of these studies were subsequently deemed to be ineligible for inclusion. One of these papers was an abstract by Yola et al (published in 2006) - it was not clear whether the patients being administered interferon-gamma had drug susceptible or drug resistant TB (it was likely that they had drug susceptible TB as the regimen was described as "standard TB medication"). (14) The other paper was a study by Suarez-Mendez et al., a frequently cited paper on the use of adjuvant interferon-gamma among patients with drug resistant TB.(11) This was an open-label, nonrandomized, non-controlled, pilot study which recruited eight patients who were given 1 000 000 IU of human recombinant interferon-gamma intramuscularly, over 24 weeks. (11) Outcomes of interest were sputum and culture conversion, lesion size on chest x-ray, clinical parameters and adverse events.(11) The authors describe that interferon-gamma was well tolerated and that sputum smears and cultures became negative before three months of treatment in all patients and that lesion size was reduced at the end of 6 months treatment.(11) Clinical improvement was also evident and few adverse events were reported.(11) However, this study did not include a comparison group, which meant that estimates of effect were not presented.(11)

An additional two studies published in Chinese journals reported on small studies of the use of interferon-gamma among patients with MDR-TB.(8, 15) The paper by Shi et al reported on the outcomes of culture conversion by six months and adverse reactions among 22 patients who were randomly assigned to intervention (interferon-gamma) or control. (15) Patients who received interferon-gamma were 2.25 more likely to convert from positive to negative by 6 months, however adverse events were slightly more likely in the interferon-gamma group (relative risk of 1.17, no confidence intervals presented).(15) The study had many limitations however, including the small sample size, the lack of information presented in the paper on randomisation or concealment. In addition the authors claimed no significant difference between the two groups in related to age and sex, yet no other baseline demographic or clinical features were reported to be adjusted for in the outcome analysis.(15) The study by Yang et al. reported on 57 patients randomised to receive interferon-gamma (n=28) or assigned to a control group (n=29). Adverse events were the main outcome of interest and there was a slightly higher risk of adverse events reported among in the intervention group (relative risk of 1.18, no confidence intervals were provided).(16) It appears that the MDR-TB regimen used in this study was a 12 month one.(16) The authors reported culture

conversion at 3 months only and smear conversion at 3, 6 or 12 months.(16) Blinding was not reported and there appeared to be no adjustment possible confounding factors in the analysis.(16)

3c. Current evidence on Perchlozone: Perchlozone is a new drug from the thiosemicarbazone drug class.(17) It was developed in Russia and is currently licensed to a Russian company, JCS Pharmasyntez.(17) It was approved in Russia for treating MDR-TB in November 2012.(17, 18) The English language literature on Perchlozone is very limited, especially with regards to original studies. One report, published in English, contains original data.(19) The data from this report was later published in a Russian language publication(20) which was one of four Russian language publications from which we extracted data. In the final summary table, we included the four Russian publications on Perchlozone; two of which were theses from Saint Petersburg State University(21, 22), the other two, publications.(20, 23) We also included adverse events data taken from the English language report by Pavlova et al.(19) The populations included in these studies were patients diagnosed with MDR or extensively drug resistant (XDR)-TB who were randomised to receive a regimen inclusive of Perchlozone or to a standard MDR or XDR regimen. The number of patients in each study ranged from 49 in the Pavlova et al. study(20) to 92 in the Belyaeva et al. study (which was the one study on XDR-TB).(22) All four studies were published between 2015 to 2018.

With regards to adverse events noted in the report by Pavlova et al., in the intervention group administered Perchlozone (n=25), 19 (76%) experienced an adverse event.(19) In the control group (n=24) this figure was 17 (70.8%). (19) The majority of adverse events were mild or moderate in nature, with 1 grade 3 adverse event recorded (in the intervention group).(19)

Six to 8 month culture conversion was reported in all four studies. Culture conversion at 8 months was reported in the one study that included XDR-TB patients.(22) No end of treatment outcomes were reported however in any of the four studies. The proportion of patients who converted from positive to negative by 6 to 8 months who received a Perchlozone containing regimen ranged from 51% (8 month culture conversion among a group XDR-TB patients) to 88.6% among 60 MDR-TB patients.(23) When comparing the intervention versus control groups, the odds ratios for the outcome of 6 to 8 month culture conversion ranged from 0.83 to 5.1 (only the Yablonskii P et al (2016) series reported statistically significant culture conversion). The outcome of culture conversion was pooled in a random effects meta-analysis with a resultant odds ratio of 1.64 (95 % confidence interval 0.76-3.52; I²=39%). The resultant GRADE evidence table and the Forest plot of the meta-analysis are presented in Appendices four and five.

### Appendix one: PICO 2 of the WHO treatment guidelines for RR and MDR-TB, 2018 update

**Q2.** In patients with rifampicin-resistant or multidrug-resistant TB (MDR/RR-TB), which individual agents when forming part of a longer regimen are more likely to improve outcomes?<sup>2</sup>

Population	Intervention	Comparator	Outcomes
MDR/RR-TB <u>without</u> resistance <u>or</u> severe intolerance to the second line drugs	A 2 <sup>nd</sup> line regin	nen that INCLUDES <sup>3</sup> :	Culture conversion by 6     months
MDR/RR-TB <u>with</u> resistance <u>or</u> severe intolerance to	<ul> <li>fluoroquinolones (Mfx/Lfx/Gfx)</li> <li>injectable agents (Km/Am/Cm)<sup>4</sup></li> <li>prothionamide or ethionamide</li> </ul>	<ul> <li>no fluoroquinolones (Mfx/Lfx/Gfx)</li> <li>no injectable agents (Km/Am/Cm)</li> <li>no prothionamide or ethionamide</li> </ul>	Successful completion of treatment     Bacteriological cure by end of
<ul> <li>Fluoroquinolones (by mutation pattern)</li> <li>2<sup>nd</sup> line injectable agents, both classes</li> <li>Fluoroquinolones + 2<sup>nd</sup> line injectable agents (i.e. XDR-TB +/- other resistance)</li> <li>Pyrazinamide</li> </ul>	- cycloserine or terizidone - linezolid - clofazimine	<ul><li>no cycloserine or terizidone</li><li>no linezolid</li><li>no clofazimine</li></ul>	<ul><li>treatment</li><li>Adherence to treatment (or treatment interruption due to</li></ul>
	<ul><li>pyrazinamide</li><li>high-dose isoniazid</li><li>ethambutol</li></ul>	- no pyrazinamide - no high-dose isoniazid - no ethambutol	non-adherence)  Treatment failure or relapse Survival (or death)
<ul> <li>Group C agents (ethionamide, prothionamide, cycloserine, terizidone, linezolid and clofazimine)</li> </ul>	- bedaquiline - delamanid	- no bedaquiline - no delamanid	Adverse reactions from anti- TB medicines     Association (complification) of
Children (0-4y, 5-14y, 10-19y), persons with HIV (on ARVs), pregnant women, persons with diabetes	- individual Group D3 agents  - sutezolid <sup>5</sup> - interferon G <sup>6</sup>	<ul> <li>no individual Group D3 agents</li> <li>no sutezolid</li> <li>no interferon G</li> </ul>	Acquisition (amplification) of drug resistance
diabetes	- perchlozone <sup>5</sup>	- no perchlozone	1

-

<sup>&</sup>lt;sup>2</sup> The evidence for this PICO question will be used to review the currently recommended cascade for longer regimen design (see Table 6 of (9)). Amongst others, the best use of bedaquiline and delamanid and their combined use will be examined.

<sup>&</sup>lt;sup>3</sup> Very few studies are known to have made head-to-head comparisons of MDR-TB medicines at different dosages. If such information is available it will be used to guide the systematic review and presentation of effects.

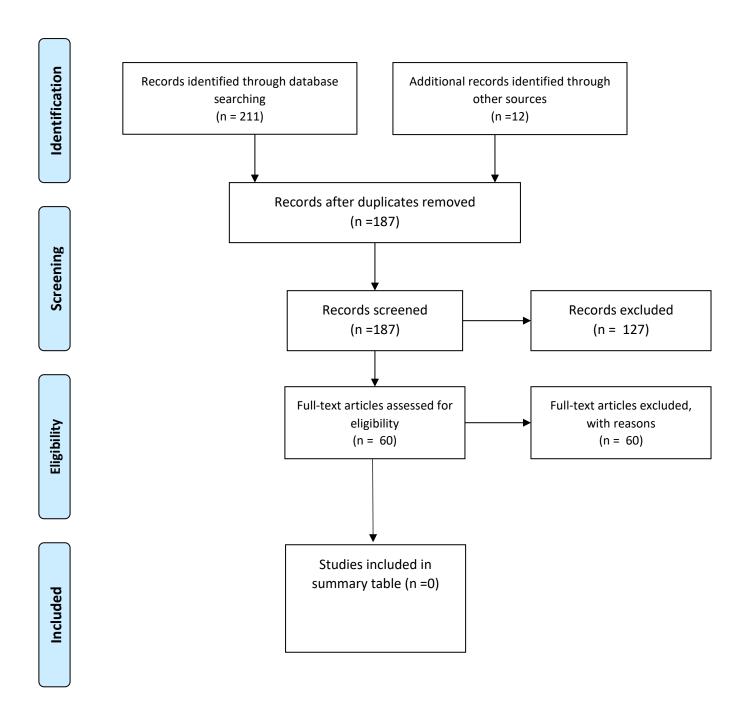
<sup>&</sup>lt;sup>4</sup> Additional detail on the effectiveness and safety of injectable agents used three times weekly (at 15mg or 25mg/kg/day) vs daily would be of interest

<sup>&</sup>lt;sup>5</sup> As yet unauthorized by stringent regulatory authorities

<sup>&</sup>lt;sup>6</sup> For a definition of « effective agents » see Chapter 5 of (8)). The evidence for this PICO question will be used to review the currently recommended cascade for longer regimen design (see Table 6 of (9))

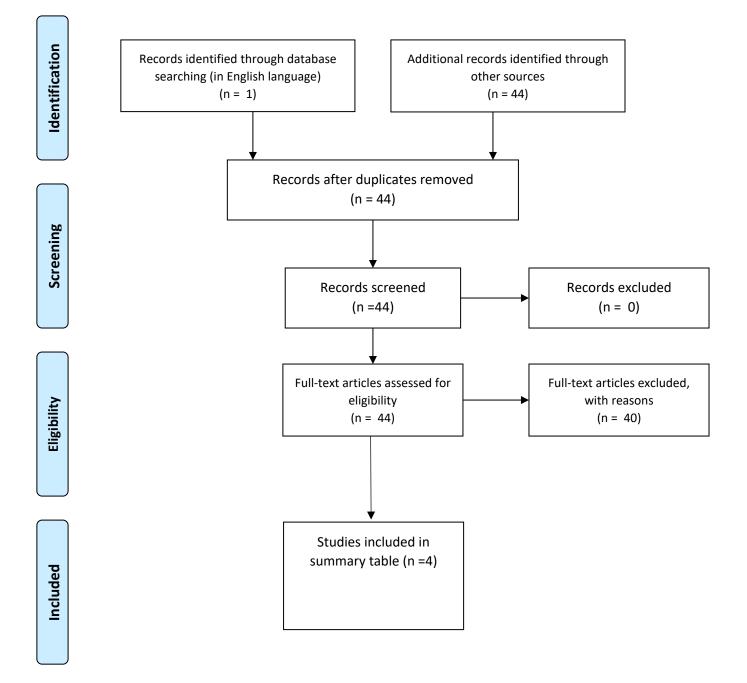
Appendix two: PRISMA flow diagrams for literature reviews on the use of Sutezolid, Interferon gamma and Perchlozone for the WHO treatment guidelines for rifampicin-resistant and multidrugresistant tuberculosis, 2018 update

# PRISMA Flow Diagram7 - SUTEZOLID



<sup>&</sup>lt;sup>7</sup> From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097. For more information, visit <a href="https://www.prisma-statement.org">www.prisma-statement.org</a>.

# **PRISMA Flow Diagram - PERCHLOZONE**



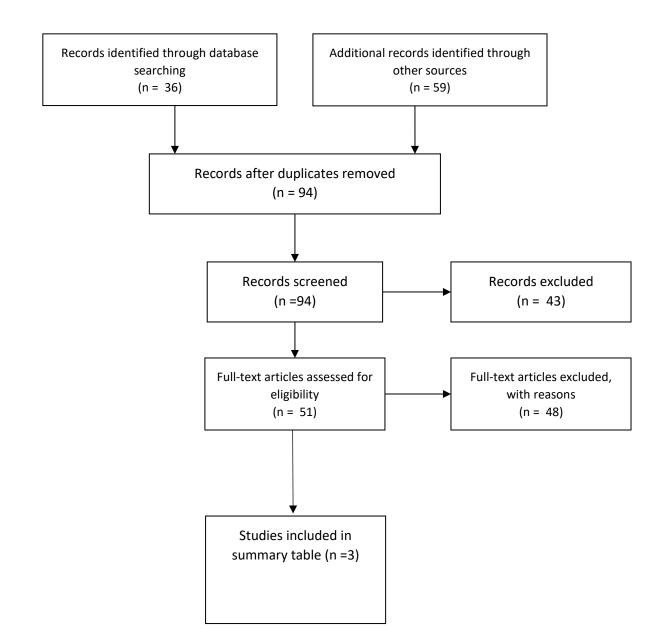
# PRISMA Flow Diagram - INTERFERON-GAMMA

Identification

Screening

Eligibility

Included



# Appendix three: Draft GRADE evidence table on the use of Perchlozone among patients with RR/ MDR-TB on the outcome of culture conversion by 6-8 months

Author(s):K Viney Date: 27 June 2018

Question: Perchlozone added to an MDR TB regimen compared to MDR TB regimen without Perchlozone in adults with MDR TB

Setting: Russian health facilities

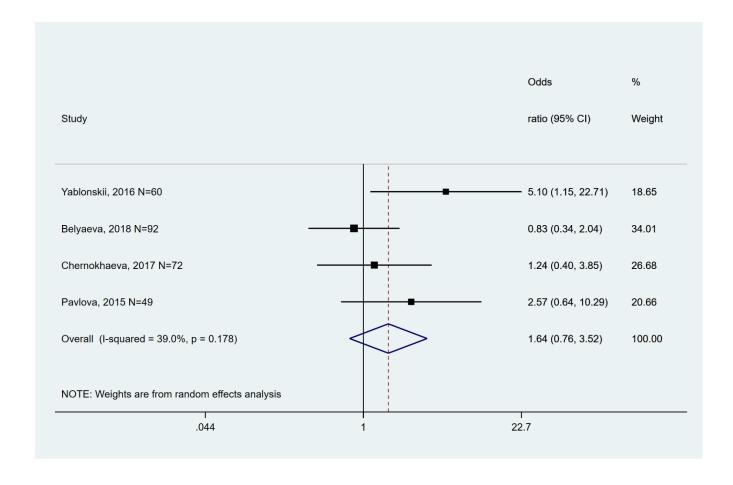
			Quality assess	ment			Nº of pa	atients	Effect			
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Perchlozone added to an MDR TB regimen	MDR TB regimen without Perchlozone	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Culture c	Culture conversion at 6 to 8 months (follow up mean 6.4 months)											
4	Randomised controlled trials	serious <sup>a</sup>	not serious <sup>b</sup>	serious <sup>c</sup>	serious <sup>d</sup>	none	100/ 144 (69.4%)	76/ 129 (58.9%)	<b>aOR 1.64</b> (0.76 to 3.52)	11 more per 100 (from 7 fewer to 25 more)	⊕○○○ VERY LOW	IMPORTANT
Adverse 6	Adverse events (follow up mean 6 months)											
1 f	Randomised controlled trial	not serious	not serious	not serious	very serious <sup>g</sup>	none	19/ 25 (76.0%)	17/ 24 (70.8%)	<b>RR 1.073</b> (0.765 to 1.505)	5 more per 100 (from 17 fewer to 36 more)	⊕○○○ VERY LOW	CRITICAL

aOR: Adjusted odds ratio; CI: Confidence interval; RR: Relative risk

#### Explanations

- a. All studies were described as randomised trials and use of Perchlozone may be associated with patient characteristics. In addition, two publications were described as theses and it is not clear how these were peer reviewed.
- b. Heterogeneity estimated using simulated I squared (39%; p=0.178).
- c. All studies were carried out in the Russian Federation and the majority were carried out in Saint Petersburg, therefore the population under study is restricted to one country.
- d. CI does not exclude an appreciable harm and all studies were small in nature (in terms of the population included in the study).
- e. Based on random effects meta-analysis pooling carried out with the advice of a biostatistician.
- f. We were only able to extract adverse events data from one trial which is a report of one of the four trials included for the outcome of culture conversion.
- q.These data were from one randomised controlled trial with small numbers in both the intervention and control arms (n=25 in the intervention arm, n=24 in the control arm). The confidence interval crosses one

# Appendix four: Forest plot from random effects meta-analysis on the use of Perchlozone among patients with RR/ MDR-TB on the outcome of culture conversion by 6-8 months



#### **References:**

- 1. Treatment Action Group. Pipeline report 2017: HIV TB HCV. New York, United States of America 2017.
- 2. Zhu T, Friedrich SO, Diacon A, Wallis RS. Population pharmacokinetic/pharmacodynamic analysis of the bactericidal activities of sutezolid (PNU-100480) and its major metabolite against intracellular Mycobacterium tuberculosis in ex vivo whole-blood cultures of patients with pulmonary tuberculosis. Antimicrobial Agents and Chemotherapy. 2014;58(6):3306-11.
- 3. Wallis RS, Jakubiec W, Kumar V, Bedarida G, Silvia A, Paige D, et al. Biomarker-assisted dose selection for safety and efficacy in early development of PNU-100480 for tuberculosis. Antimicrobial Agents and Chemotherapy. 2011;55(2):567-74.
- 4. Wallis RS, Jakubiec WM, Kumar V, Silvia AM, Paige D, Dimitrova D, et al. Pharmacokinetics and whole-blood bactericidal activity against Mycobacterium tuberculosis of single doses of PNU-100480 in healthy volunteers. The Journal of Infectious Diseases. 2010;202(5):745-51.
- 5. Wallis RS, Jakubiec W, Mitton-Fry M, Ladutko L, Campbell S, Paige D, et al. Rapid evaluation in whole blood culture of regimens for XDR-TB containing PNU-100480 (sutezolid), TMC207, PA-824, SQ109, and pyrazinamide. PLoS One. 2012;7(1):e30479.
- 6. Wallis RS, Dawson R, Friedrich SO, Venter A, Paige D, Zhu T, et al. Mycobactericidal activity of sutezolid (PNU-100480) in sputum (EBA) and blood (WBA) of patients with pulmonary tuberculosis. PLoS One. 2014;9(4):e94462.
- 7. Condos R, Rom WN, Schluger NW. Treatment of multidrug-resistant pulmonary tuberculosis with interferon-γ via aerosol. The Lancet. 1997;349(9064):1513-5.
- 8. Dawson R, Condos R, Tse D, Huie ML, Ress S, Tseng C-H, et al. Immunomodulation with recombinant interferon-γ1b in pulmonary tuberculosis. PloS One. 2009;4(9):e6984.
- 9. Park S-K, Cho S, Lee I-H, Jeon D-S, Hong S-H, Smego RA, et al. Subcutaneously administered interferon-gamma for the treatment of multidrug-resistant pulmonary tuberculosis. International Journal of Infectious Diseases. 2007;11(5):434-40.
- 10. Koh W-J, Kwon OJ, Suh GY, Chung MP, Kim H, Lee NY, et al. Six-month therapy with aerosolized interferon-γ for refractory multidrug-resistant pulmonary tuberculosis. Journal of Korean Medical Science. 2004;19(2):167-71.
- 11. Suárez-Méndez R, García-García I, Fernández-Olivera N, Valdés-Quintana M, Milanés-Virelles MT, Carbonell D, et al. Adjuvant interferon gamma in patients with drug—resistant pulmonary tuberculosis: a pilot study. BMC Infectious Diseases. 2004;4(1):44.
- 12. Wallis RS. Lack of a Therapeutic Role for Interferon γ in Patients With Tuberculosis. The Journal of Infectious Diseases. 2014;209(4):627-8.
- 13. Business Wire. InterMune enrolls first patient in phase III trial in multidrug-resistant tuberculosis 2000 California, United States of America Business Wire; 2000 [Available from: <a href="https://www.thefreelibrary.com/InterMune+Enrolls+First+Patient+in+Phase+III+Trial+in...-a063778269">https://www.thefreelibrary.com/InterMune+Enrolls+First+Patient+in+Phase+III+Trial+in...-a063778269</a>.
- 14. Yola A, Sologub T, Nechaev V, Ivanov A. Immune-based therapy using gamma interferon ingaron in the treatment of HIV/AIDS patients with active pulmonary tuberculosis (PTB) not previously highly active antiretroviral therapy (HAART). Retrovirology. 2006;3(S1):S38.
- 15. Shi JW, Yuan Y, Gu HY. Effect of aerosolised interferon gamma in 11 patients with multi drugresistant tuberculosis Journal of Clinical and Pulmonary Medicine. 2008;13:1653-4.
- 16. Yang Z, Wang L, Guo L, Lin M, Chen X. A short erm efficacy of aerosolising inhalation recombinant human gamma interferon in multidrug resistant pulmonary tuberculosis Journal of the Chinese Tuberculosis Association. 2009;3(11):660-3.
- 17. Gopal P, Dick T. The new tuberculosis drug Perchlozone® shows cross-resistance with thiacetazone. Int J Antimicrob Agents. 2015;45(4):430-3.
- 18. National Association of Phthisiatricians. Clinical Recommendations on Diagnostics and Treatment of Respiratory Tuberculosis in Adults. Russian Federation National Association of Phthisiatricians; 2013.

- 19. Pavlova M, Starshinova A, Sapoznhikova N, Chernokhaeva I, Yakovchuk A, Archakova L, et al. Efficacy of complex therapy and possible adverse reactions of respiratory apparatus multi-drug resistance tuberculosis treatment. Saint Petersburg, Russian Federation Saint Petersburg State University 2014.
- 20. Pavlova M, Vinogradova T, Starshinova A, Sappozhnikova N, Chernokhaeva I, Archakova L, et al. Confirmation of preclinical studies data on Perchlorate (Perchlozon) in clinic. International Journal of Current Research 2015;7(4):14501-5.
- 21. Chernokhaeva I. The effectiveness tioureidoiminomethylpyridiunium perchlorate in the treatment of multidrug-resistant tuberculosis. Saint Petersburg, Russian Federation Saint Petersburg State University 2017.
- 22. Belyaeva E. Features of the course and effectiveness of therapy for extensively drug resistant pulmonary tuberculosis Saint Petersburg, Russian Federation Saint Petersburg State University 2018.
- 23. Yablonskii P, Vinogradova T, Levashev Y, Pavlova M, Zilber E, Starshinova A, et al. Preclinical and clinical trials of the new tuberculosis drug perchlozon. Therapeutic Archive. 2016;88(3):111-5.

#### **Bedaquiline**

#### Source Data

Pediatric data for bedaquiline were reviewed to explore the extent to which adult data can be extrapolated to children. The focus of this review was on safety and pharmacologic exposure data available from 2 ongoing pediatric studies:

- TMC207-C211: A Phase 2, Open-label, Multicenter, Single-arm Study to Evaluate the Pharmacokinetics, Safety,
  Tolerability and Anti-mycobacterial Activity of TMC207 in Combination With a Background Regimen (BR) of
  Multidrug Resistant Tuberculosis (MDR-TB) Medications for the Treatment of Children and Adolescents 0 Months to
  <18 Years of Age Who Have Confirmed or Probable Pulmonary MDR-TB.</li>
- 2. IMPAACT P1108: A Phase I/II, Open-Label, Single Arm Study to Evaluate the Pharmacokinetics, Safety and Tolerability, of Bedaquiline (BDQ) in Combination with optimized Individualized Multidrug-Resistant Tuberculosis (MDR-TB) on in HIV-Infected and HIV-Uninfected Infants, Children and Adolescents with MDR-TB Disease.

The data available from TMC207-C211 included a sponsor generated summary document for patients from cohort 1 along with suppressed raw data files for all enrolled participants in the 12-18 yr cohort. From IMPAACT P1108, summary data were provided for participants enrolled to date with selected raw data provided on follow-up request. For the purposes of comparison, adult reference data were extracted from the publically accessible FDA review of application #204384. With respect to limitations, it was noted that both trials are still ongoing and not all of the data provided were verifiable. However, where possible, sponsor summary presentations were examined against the sponsor raw data and sponsor derived pharmacokinetic exposure parameters were examined against independently calculated pharmacokinetics.

#### **Patient Characteristics**

Enrolled Participants	TMC207-C211	IMPAACT P1108
Count	15 <sup>a</sup>	10 <sup>b</sup>
Nationality/Race	Philippines (2): Russia (5): South Africa (8)	Black, non-Hispanic (10)
Comorbid HIV	excluded	included <sup>c</sup>
Age (yr)	14 - 17	6 - 17
Sex (M:F)	3:12	6:4
Height (cm)	150 – 175	121 - 162
Weight (kg)	38.4 – 75	15 - 65
BMI z-score	-2.8 to 1.7	N/A
Dose <sup>d</sup> : cohort 1 : ≥ 30 kg	400 mg QD 2wk; 200 mg TIW 22 wk	400 mg QD 2wk; 200 mg TIW 22 wk
N/A : ≥ 15 to <30 kg		200 mg QD 2wk; 100 mg TIW 22 wk

<sup>&</sup>lt;sup>a</sup> only 6 participants with 2 week PK data

#### Safety

Physical exam/Laboratory based safety. As of the meeting date, 14 children from cohort 1 experienced adverse events (AE) in TMC207-C211. Those occurring at a rate of greater than 10% included hypoacusis, tinnitus, eye pain, blurred vision, nausea, URI, vulvovaginal candidiasis, prolonged prothrombin time (PT), arthralgia, acne, and rash. Among these, the only AE for which a safety signal had net been previously reported in adults was elevated PT. There were 5 grade 3 or 4 AE: 1 episode of elevated ALT/AST/Bili which did not recur upon de/rechallenge; 1 episode of elevated CPK which resolved after 1 day; and 3 episodes of elevated PT two of which were preceded by an elevated PT at baseline or screening, none of which appeared to be the result of an underlying coagulation disorder, and all of which resolved at next scheduled visit (4wk). The investigators of IMPAACT P1108 reported no grade 3 or 4 physical exam or laboratory based safety endpoints.

<sup>&</sup>lt;sup>b</sup> only 9 participants with safety data and 7 participants with week 24 PK data

<sup>&</sup>lt;sup>c</sup> no HIV positive participants enrolled to date

d IMPAACT P1108 used a body weight of <30 kg as the threshold to adjust the paediatric dose to half the adult dose whereas TMC207-C211 used age <12 years to make the same dose adjustment.

Cardiac safety. In TMC207-C211, 6 participants experienced an increase in QTcF of 30-60 msec. In IMPAACT P1108, 1 participant experienced an increase in QTcF of 93.7 msec. Concurrent QTc prolonging medications included levofloxacin in both studies and clofazamine on the IMPAACT trial.

#### Exposure

Observed PK data. Observed 2- and 12-week Cmax, Cmin and AUC<sub>0-24hr</sub> data in the TMC207-C211 pediatric cohort were available for comparison with the corresponding 2- and 24-week exposure data from adults (Figure 1). The expected markers of internal consistency (i.e. dose-exposure and exposure-exposure relationships were present) were observed in the TMC207-C211 data set. In addition, the reported covariate relationship in adults, specifically the race effect, appeared to be reproducible in the adolescent cohort.

Population PK data. AUC<sub>0-168hr</sub> data predicted using a popPK approach were available from both pediatric studies along with the reference adult trial (Figure 2). The age subgroups represented were selected to parallel the actual age range of children enrolled in TMC207-C211, the intended age range for cohort 1 in TMC207-C211, and the children expected in cohort 2 of TMC207- C211.

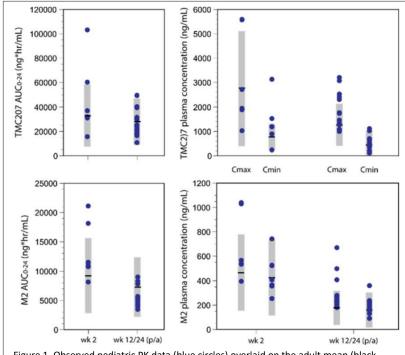


Figure 1. Observed pediatric PK data (blue circles) overlaid on the adult mean (black bar) ± 2 standard deviations (grey boxes).

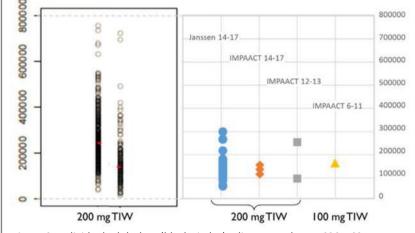


Figure 2. Individual adult data (black circles) adjacent to the TMC207-C211 adolescents (blue circles) and the IMPAACT P1108 participants 14-17 yr (orange diamonds), 12-13 yr (grey squares) and 6-11 yr (yellow triangles).

It should be highlighted that the popPK models used by the Janssen and IMPAACT study teams are not identical and the extent to which they differ is unclear. Visual predictive checks for both models were available for GDG review.

# Considerations by the GDG in their recommendation

Will the administration of BDQ harm adolescents to whom it is administered? The GDG preliminarily concluded that the safety risk in children down to 6 years of age and constituted by the population enrolled in these trials (e.g. HIV negative, limited exposure to concomitant medications with the potential to prolong QTc, etc.) does not appear to exceed that of adults.

Will the administration of BDQ provide benefit for adolescents to whom it is administered? Operating under the assumption that exposure-response (efficacy) profiles can be extrapolated from adults to children, the GDG preliminarily concluded that the doses evaluated do not appear to produce exposures that would put children at increased risk for therapeutic failure. Note that the variability present in the limited sample size precluded a comment on exposure- response (safety).

Can the risk:benefit balance be shifted in favor of benefit? The GDG preliminarily concluded that risk-benefit considerations for the use of bedaquiline are similar to those considered for adults. The GDG also endorsed the need to amass additional data in children before the strength of any recommendation can be modified.

#### Delamanid

#### Source Data

Pediatric data for delamanid were reviewed to examine whether the recommendations for delamanid use in children can be further lowered to children under 6 years of age (from 12 yr in the 2016 guidance document). The focus of this summary was on safety and pharmacologic exposure data available from ongoing pediatric studies with the following data available:

1. 242-12-232: Open-label, age de-escalation trial designed to assess the pharmacokinetics, safety, and tolerability of delamanid administered twice daily for 10 days in children with multidrug-resistant (MDR) tuberculosis on therapy with an optimized background regimen (OBR).

raw data for cohorts 1 & 2 last updated May 2016 raw data for cohorts 3 & 4 provided Aug 2018

2. 242-12-233: Ongoing phase 2, open-label, multi-dose trial to assess the safety, tolerability, PK, and efficacy of delamanid in children with MDR-TB on Therapy with an OBR over a 6-month treatment period.

raw data for cohorts 1 & 2 last updated May 2016 raw data for cohorts 3 provided Aug 2018

- 3. 242-07-204: selected raw and summary data adult data from this phase 2 study of delamanid administered twice-daily for 8 weeks in addition to OBR in subjects with MDR-TB were available for comparison purposes. Where necessary, data were extracted from EMA Assessment report EMEA/H/C/002552.
- 4. 242-12-245: A Phase 1, Randomized, Open-label, Single-dose, Two-way Crossover, Relative Bioavailability Study Comparing a 100-mg Oral Dose of Delamanid Tablets and a 100 mg Oral Dose of the Delamanid Pediatric Formulation in Healthy Adult Subjects

full clinical study report

additional personal communications related the formulation were provided by Otsuka

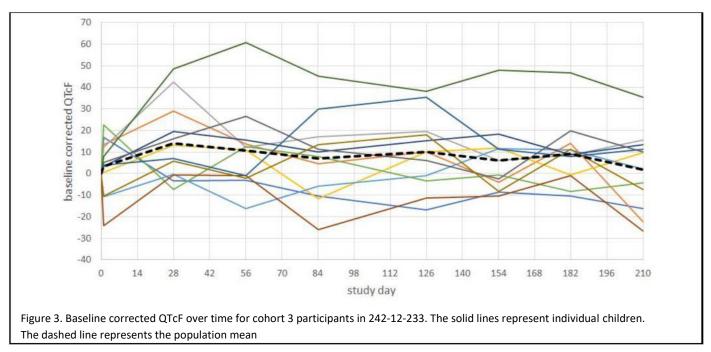
The pharmacokinetic data provided were restricted to delamanid and a singular metabolite (M6705). Sponsor PK data were examined against independently analyzed PK with 98% concordance in calculated PK parameters between sponsor and reviewer. Sponsor provided QTc data (in triplicate) were averaged at each time point with the QTc data collected at baseline used to determine QTc during the treatment phase. Both Bazett and Fredericia corrected QTc interval data were provided for review. The latter (QTcF) is discussed in the section below owing to its increased accuracy in the presence of elevated RR intervals (often seen in children) and to maintain consistency in comparison with previously evaluated pediatric and adult data.

### **Patient Characteristics**

	5 mg QD (<3 yr)	5 mg BID (<3 yr)	10 mg BID (<3 yr)	25 mg BID (3-5 yr)	50 mg BID (6-11 yr)	100 mg BID (12-17 yr)
cohort	4	4	4	3	2	1
count	2	4	6	12	6	7
age (yr)	0.7, 1.1	1.8 ± 0.7	1.8 ± 0.4	4.3 ± 1.0	9.2 ± 1.5	14.7 ±1.5
sex (M:F)	1:1	0:4	5:1	6:6	2:4	4:3
ethnicity (As:Afr)	1:1	1:3	4:2	8:4	4:2	7:0
height (cm)	68,63	79.0 ± 4.5	82.2 ± 5.6	97.4 ± 8.6	122.3 ±11.3	150.9 ±10.8
weight (kg)	8.0, 5.5	9.5 ± 0.5	11.0 ± 1.1	14.2 ± 3.2	24.9 ±6.8	38.0 ±6.4
BMI (kg/m <sup>2</sup> )	17.3, 13.9	15.2 ± 1.2	16.4 ± 2.7	14.8 ± 1.4	16.3 ±2.4	16.7 ±2.2
zBMI/W4L	-0.09, -2.58	-1.15 ± 0.8	-0.47 ± 1.9	-0.81 ± 1.1	-0.25 ± 0.9	-1.6 ± 1.2
daily dose (mg/kg)	0.63, 0.91	1.1 ± 0.1	1.8 ± 0.2	3.7 ± 0.8	4.3 ± 1.3	5.4 ± 1.1

#### Safety

Cardiac Safety. Across the entire cohort, a weak but insignificant (p=0.066) temporal relationship between delamanid administration and QTcF was observed (Figure 3). Between days 28 and 182, the intercept of QTcF vs. time was 8.57 milliseconds; however, the 95% confidence included zero (95% CI -0.57 to 17.71). Three of the 12 children enrolled in cohort 3 experienced an increase in QTcF of greater than 30 milliseconds and one of these children experienced an in increase in excess of 60 milliseconds. In 2 children the elevations were isolated to 1 or 2 observations. The third child experienced a sustained elevation in QTcF over the duration of the treatment period as can be seen in Figure 3.



Laboratory based safety. Across cohort 3; chemistry, hematology, coagulation, lipids, and urinalysis testing reveal limited changes during the treatment period. The only laboratory parameters reaching statistical significance for a temporal relationship with protracted medication administration were a reduction in calcium (P<0.01) and uric acid (p=0.013), and an increase in absolute neutrophil count (p=0.048); however, the clinical consequence of these changes is limited. With respect to laboratory abnormalities in individual children, the majority reflected isolated events.

When looking specifically at repeated or sustained abnormalities which occurred during the treatment period and were not present at screening or baseline; 1 child experienced an elevation in AST, ALT and GGT starting on day 126 and continuing through day 138 for the transaminases and day 189 got GGT.

#### Exposure

There were appreciable differences in weight adjusted total daily dose (i.e. mg/kg/day) between cohorts (Figure 4). Only children in cohorts 1-3 received doses comparable to those observed in the adult trial. Consequently, exposures at all doses explored in cohort 4 fell below adult values (Figure 5).

Z-scores for body-mass-index (weight-for length in children under 2 years) was also unevenly distributed between the dosing cohorts (Figure 6, left). The combined result is a confounding of dose- exposure by both age and nutritional status. As a result non-monotonic relationships between age and dose-corrected delamanid exposure were observed with the first and last dose of

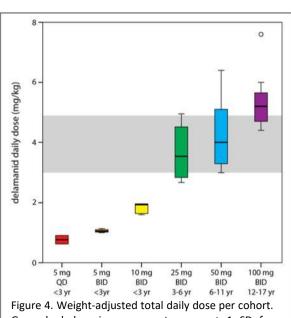
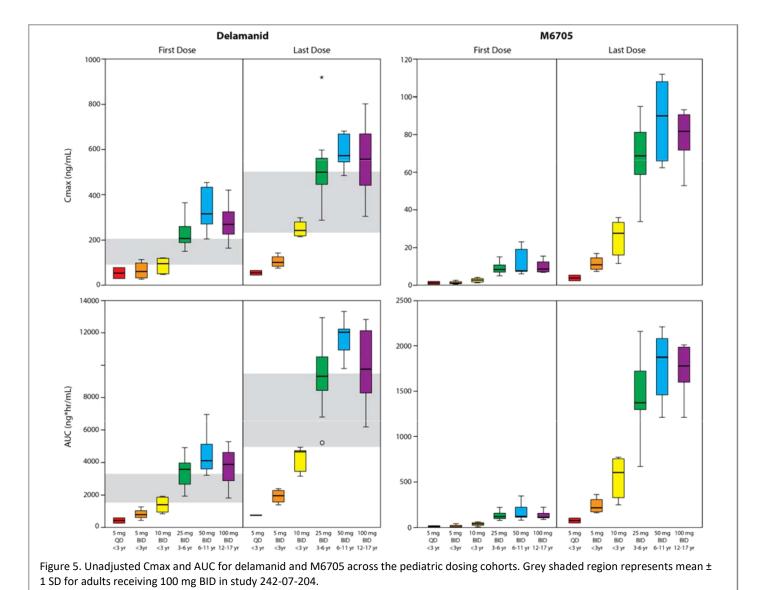
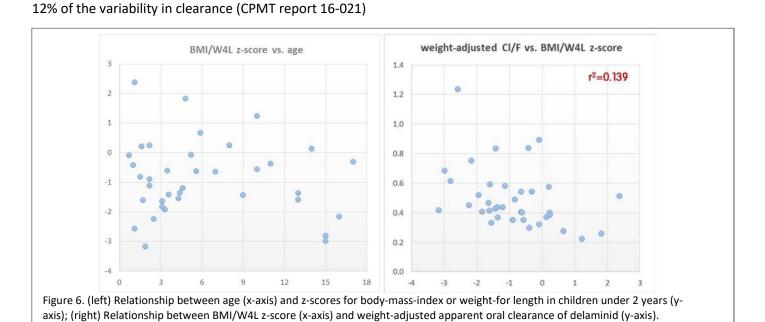


Figure 4. Weight-adjusted total daily dose per cohort. Grey shaded region represents mean ± 1 SD for adults receiving 100 mg BID in study 242-07-204



the PK study. Notably, with the increase in sample size, the relationship between nutritional status and delamanid clearance are now more consistent between children (Figure 6, right) and adults in whom nutritional status accounts for



#### Considerations for the GDG as they prepare their final recommendation

1. Extrapolation to Children under the age of 6 years:

Extrapolations of efficacy and safety should be restricted to children 3 years of age and older.

Based on the PK data provided, exposure profiles in children 3-5 years of age were comparable to adults and no higher than in children 6 years of age and older for whom the GDG has concluded administration of delamanid is reasonable.

Based on the laboratory and cardiac data provided, no safety signals distinct from those reported in adults were observed in children 3-5 years of age.

Given non-monotonic nature of the relationship between age and exposure (Figure 6, left), estimating alternative doses to those tested in trial 242-12-232 for children under the age of 3 years is not recommended at this time.

# 2. Extemporaneous Modification of the Commercially Available Tablet:

Cohorts 3 and 4 in trial 242-12-232 were administered a scored-dispersible pediatric formulation that is <u>not</u> <u>expected to be available</u> at the time of GDG publication (Otsuka, personal communication). As a result, the only source of delamanid is the adult formulation which poses the following problems when considered for children requiring <50 mg:

The delamanid adult formulation and pediatric formulation are **not bioequivalent**. In a crossover bioequivalence (BE) study, neither Cmax [90%CI GMR 0.701,0.809] nor AUC [90%CI GMR 0.775,0.909] satisfied the criteria for BE as specified by regulatory agencies. As such, the formulations are not interchangeable. Substituting the adult formulation for the pediatric formulation will result in higher delamanid exposures than would be expected from the pediatric formulation.

The commercially available film coated tablets that were used in the adult studies employ a specific manufacturing process to create a powder blend with enhanced solubility and improved bioavailability over the drug substance (DS) in crystal form which is practically insoluble. The impact of added compression, resulting from tablet crushing, on the physical properties of the DS is unclear and could appreciably alter (most likely reduce) the bioavailability of delamanid.

The adult tablet is **not scored** and will likely shatter when attempting to split.

The spray dried material comprising the tablet is **exceedingly bitter** and will likely be unpalatable.

Delamanid is <u>susceptible to oxidation</u>. As a consequence, oxygen is purged from the blister packs during the manufacturing process. Retaining pill fragments for use at any time other than the time of administration will likely result in the delivery of lower than expected active compound and unspecified oxidation byproducts.

Temperature- and photostability have also been examined for production-scale tablets and the powder blend. Detailed or summary findings are not provided in EMEA/H/C/002552; however, sponsor confirms that delamanid is **heat sensitive** further compromising the fidelity of retained tablet fragments.

Should the GDG conclude that there is enough data to support the use of delamanid down to 3 years of age, some variation of the following draft language, provided for consideration by the GDG, should be included in the guidance document:

Important: The adult delamanid formulation has not been tested in pediatric subjects < 6 years of age. The adult and pediatric formulations of delamanid are not interchangeable. Equal doses of each formulation achieve different concentrations in the body. In addition, splitting or crushing of the adult tablet for administration to children will affect the stability of the medicine and result in pill fragments that are exceedingly bitter.