Heterogeneity and subgroup analyses

The wide credibility intervals found in our review point to the great heterogeneity of the results presented and reduce the validity of possible subgroup analyses. In addition, the clinical differences between the studies, such as the different ARV regimens used, the initial CD4 level of the participants and their previous exposure to ART, in addition to the different data analyses (by protocol or intention to treat) and funding sources (only by the industry, only by government agencies or both), mean that variations in the estimates found in the subgroup analyses may be due only to residual confounding, without any implication for clinical practice. Most of these analyses did not demonstrate differences between the groups and the few different estimates found seem to reflect the absence or very low number of events (as in the case of death) and, mainly, the clinical heterogeneity, especially the different ARV regimens evaluated in the studies.

Subgroup analyses for the primary endpoint results of this review, relating to blinding, per-protocol or intention-to-treat analysis, CD4 level, prior ART exposure, and funding, are shown in Table S7. Data regarding CD4 level and funding were not presented for total clinical and/or laboratory adverse events and clinical and/or laboratory adverse events related to ART, as they did not have subgroups. Both outcomes were only reported in studies with participants whose average baseline CD4 level was greater than 250 cells/mm^3 and with exclusive funding from the pharmaceutical industry.

We did not perform a subgroup analysis by ARV regimen, as the same regimen was evaluated in a maximum of two arms of the included studies. We also did not perform some of the analyses provided in our protocol, due to available data. The studies had adult women, with very few elderly women, and there were no children or adolescents, which did not allow analysis by age group. It was also not possible to evaluate according to the level of development of the study sites, since 80% of them were multicenter and were conducted in countries with different levels of development, as defined by the World Bank. The follow-up time was used to adjust the incidence rate, not allowing a subgroup analysis. In general, the studies monitored adverse events more actively and all used ARV regimens with three or more associated drugs, not characterizing the existence of different subgroups in these aspects. We did not perform subgroup analysis according to the risk of bias, as 90% of the studies were at high risk, and also according to allocation concealment, since this was directly related to blinding, as 80% of the

studies were open and the 20% that were double-blind did not describe how the concealment took place.

Table S7 - Subgroup analyses (to be continued)

Blinding analysis (double-blind vs open)							
Outcome	Subgroup	Number of study arms	Number of women	Incidence rate (number of events/1000 person-years)	Credibility interval (95%)		
Discontinuation due to adverse events	Double-blind	4		52,77	8,89	-	285,10
	Open	11		12,04	1,43	-	57,06
Discontinuation due to adverse events related to ART	Double-blind	2		76,47	1,65	-	2755,00
	Open	5		0,82	0,01	-	22,93
Clinical and/or laboratory adverse events	Double-blind	2		950,80	55,35	-	16180,00
	Open	4		874,20	629,30	-	1213,00
Clinical and/or laboratory adverse events related to ART	Double-blind	2		578,70	30,71	-	10720,00
	Open	4		264,90	46,68	-	1457,00
Analysis by intention to treat (protocol vs intention to treat)							
Outcome	Subgroup	Number of study arms	Number of women	Incidence rate (number of events/1000 person-years)	Credibility interval (95%)		
Discontinuation due to adverse events	Protocol	10		13,09	1,90	-	55,53
	Intention to treat	5		53,61	3,90	-	316,20
Discontinuation due to adverse events related to ART	Protocol	4		1,34	0,01	-	49,34
	Intention to treat	3		20,77	0,18	-	691,30
Clinical and/or laboratory adverse events	Protocol	4		874,20	628,00	-	1213,00
	Intention to treat	2		950,80	55,35	-	16180,00
Clinical and/or laboratory adverse events related to ART	Protocol	4		264,70	46,89	-	1468,00
	Intention to treat	2		579,00	30,71	-	10710,00
Analysis by CD4 level (mean >250 cells/mm^3 vs mean <250 cells/mm^3)							
Outcome	Subgroup	Number of study arms	Number of women	Incidence rate (number of events/1000 person-years)	Credibility interval (95%)		
Discontinuation due to adverse events	>250	11		33,45	8,98	-	87,29
	<250	4		5,44	0,10	-	142,70
Discontinuation due to adverse events related to ART	>250	5		6,26	0,12	-	106,40
	<250	2		1,98	0,00	-	413,39

Table S7 - Subgroup analyses (conclusion)

Analysis by previous exposure to ART (treatment naïve vs experienced)							
Outcome	Subgroup	Number of study arms	Number of women	Incidence rate (number of events/1000 person-years)	Credibility interval (95%)		
Discontinuation due to adverse events	Treatment naïve	12		25,53	6,36	-	77,30
	Experienced	3		5,95	0,04	-	282,49
Discontinuation due to adverse events related to ART	Treatment naïve	5		7,53	0,14	-	161,60
	Experienced	2		0,97	0,00	-	85,97
Clinical and/or laboratory adverse events	Treatment naïve	4		952,90	781,20	-	1153,00
	Experienced	2		794,40	42,10	-	15130,00
Clinical and/or laboratory adverse events related to ART	Treatment naïve	4		520,00	305,40	-	909,49
	Experienced	2		140,10	3,73	-	5242,95
Analysis by funding (industry vs mixed[1] vs government)							
Outcome	Subgroup	Number of study arms	Number of women	Incidence rate (number of events/1000 person-years)	Credibility interval (95%)		
Discontinuation due to adverse events	Industry	9		36,10	9,99	-	98,27
	Mixed[1]	4		5,49	0,10	-	142,60
	Governmental	2		10,22	0,01	-	1559,95
Discontinuation due to adverse events related to ART	Industry	4		11,74	0,24	-	256,50
	Mixed[1]	2		1,98	0,00	-	417,50
	Governmental	1		0,00	0,00	-	0,48

1. Mixed: study funded by industry and government agencies

Assessing the certainty of the evidence

Grading of Recommendations, Assessment, Development and Evaluation (GRADE) is normally used to assess the quality of evidence in studies that make comparative analyses between different interventions. Our review is descriptive, estimating the incidence rate of events per 1000 person-years for each intervention, using the arms of the primary studies as the unit of analysis. In any case, we chose to elaborate GRADE, using its tool for evaluating the certainty of the evidence for prognostic questions (Table S8).

Table S8 - Summary of findings (to be continued)

№ of studies	Certainty assessment						Effect			Certainty	Importance
	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	№ of events	№ of individuals	Rate (95% CrI)		
Discontinuation due to adverse events (follow-up: range from 48 weeks to 168 weeks; assessed with: incidence rate)											
9[1,2,3,4,5,6,7,8,9]	randomized trials	very serious[a]	very serious[b]	not serious	very serious[c]	None	131	2388	event rate 20.78 per 1000 person-years (5.58 to 57.31)	⨁◯◯◯ Very low	IMPORTANT
Discontinuation due to adverse events related to ART (follow-up: range from 48 weeks to 168 weeks; assessed with: incidence rate)											
5[2,3,4,5,6]	randomized trials	very serious[d]	very serious[b]	not serious	very serious[c]	none	43	988	event rate 4.31 per 1000 person-years (0.13 to 54.72)	⨁◯◯◯ Very low	IMPORTANT

Table S8 - Summary of findings (continuation)

№ of studies	Certainty assessment						Effect			Certainty	Importance
	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	№ of events	№ of individuals	Rate (95% CrI)		
Clinical and/or laboratory adverse events (follow-up: range from 48 weeks to 48 weeks; assessed with: incidence rate)											
4[3,4,5,8]	randomized trials	very serious[e]	very serious[b]	not serious	very serious[c]	none	736	954	event rate 888.20 per 1000 person-years (759.9 per 1045)	⨁◯◯◯ Very low	IMPORTANT
Laboratory adverse events (follow-up: range from 184 weeks to 184 weeks; assessed with: incidence rate)											
1[10]	randomized trials	very serious[f]	not serious	not serious	very serious[c]	none	91	483	event rate 52.02 per 1000 person-years (1.89 to 1388)	⨁◯◯◯ Very low	NOT IMPORTANT
Clinical and/or laboratory adverse events related to ART (follow-up: range from 48 weeks to 48 weeks; assessed with: incidence rate)											
4[3,4,6,8]	randomized trials	very serious[e]	very serious[b]	not serious	very serious[c]	none	298	954	event rate 341.60 per 1000 person-years (133.6 to 862.7)	⨁◯◯◯ Very low	IMPORTANT

Table S8 - Summary of findings (continuation)

№ of studies	Certainty assessment						Effect			Certainty	Importance
	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	№ of events	№ of individuals	Rate (95% CrI)		
Clinical adverse events related to ART (follow-up: range from 48 weeks to 48 weeks; assessed with: incidence rate)											
1[9]	randomized trials	very serious[f]	not serious	not serious	very serious[c]	none	221	575	event rate 431.50 per 1000 person-years (13.84 to 13630)	⨁◯◯◯ Very low	IMPORTANT
Grade 3 and/or 4 clinical and/or laboratory adverse events (follow-up: range from 48 weeks to 156 weeks; assessed with: incidence rate)											
3[3,5,6]	randomized trials	very serious[g]	very serious[b]	not serious	very serious[c]	none	37	284	event rate 96.34 per 1000 person-years (55.04 to 158.9)	⨁◯◯◯ Very low	CRITICAL
Grade 3 and/or 4 clinical adverse events (follow-up: range from 48 weeks to 168 weeks; assessed with: incidence rate)											
3[1,2,9]	randomized trials	very serious[h]	very serious[b]	not serious	very serious[c]	none	158	1316	event rate 59.93 per 1000 person-years (33.74 to 104.6)	⨁◯◯◯ Very low	CRITICAL

Table S8 - Summary of findings (continuation)

№ of studies	Certainty assessment						Effect			Certainty	Importance
	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	№ of events	№ of individuals	Rate (95% CrI)		
Grade 3 and/or 4 laboratory adverse events (follow-up: range from 48 weeks to 168 weeks; assessed with: incidence rate)											
4[1,2,4,9]	randomized trials	very serious[i]	very serious[b]	not serious	very serious[c]	none	452	1550	event rate 145.10 per 1000 person-years (57.71 to 359.9)	⨁◯◯◯ Very low	CRITICAL
Grade 3 clinical and/or laboratory adverse events (follow-up: range from 48 weeks to 48 weeks; assessed with: incidence rate)											
1[8]	randomized trials	very serious[f]	not serious	not serious	very serious[c]	none	55	495	event rate 125.50 per 1000 person-years (3.51 to 4295)	⨁◯◯◯ Very low	IMPORTANT
Grade 4 clinical and/or laboratory adverse events (follow-up: range from 48 weeks to 48 weeks; assessed with: incidence rate)											
1[8]	randomized trials	very serious[f]	not serious	not serious	very serious[c]	none	12	495	event rate 25.05 per 1000 person-years (0.5 to 1017)	⨁◯◯◯ Very low	CRITICAL

Table S8 - Summary of findings (continuation)

№ of studies	Certainty assessment						Effect			Certainty	Importance
	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	№ of events	№ of individuals	Rate (95% CrI)		
Grade 4 clinical adverse events (follow-up: range from 168 weeks to 168 weeks; assessed with: incidence rate)											
1[2]	randomized trials	very serious[f]	not serious	not serious	very serious[c]	none	15	500	event rate 9.31 per 1000 person-years (0.39 to 216.3)	⨁◯◯◯ Very low	CRITICAL
Serious clinical and/or laboratory adverse events (follow-up: range from 48 weeks to 168 weeks; assessed with: incidence rate)											
8[1,2,3,4,5,6,7,8]	randomized trials	very serious[j]	very serious[b]	not serious	very serious[c]	none	140	1813	event rate 49.34 per 1000 person-years (31.6 to 77.1)	⨁◯◯◯ Very low	CRITICAL
Grade 3 and/or 4 clinical and/or laboratory adverse events related to ART (follow-up: range from 48 weeks to 48 weeks; assessed with: incidence rate)											
2[3,6]	randomized trials	very serious[k]	very serious[b]	not serious	very serious[c]	none	6	225	event rate 27.75 per 1000 person-years (2.56 to 272.9)	⨁◯◯◯ Very low	CRITICAL

Table S8 - Summary of findings (continuation)

№ of studies	Certainty assessment						Effect			Certainty	Importance
	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	№ of events	№ of individuals	Rate (95% CrI)		
Serious clinical and/or laboratory adverse events related to ART (follow-up: range from 48 weeks to 48 weeks; assessed with: incidence rate)											
4[3,4,6,8]	randomized trials	very serious[e]	very serious[b]	not serious	very serious[c]	none	6	954	event rate 1.09 per 1000 person-years (0.01 to 21.47)	⨁◯◯◯ Very low	CRITICAL
Serious clinical adverse events related to ART (follow-up: range from 48 weeks to 48 weeks; assessed with: incidence rate)											
4[4,6,8,9]	randomized trials	very serious[i]	very serious[b]	not serious	very serious[c]	none	8	1197	event rate 1.53 per 1000 person-years (0.01 to 21.29)	⨁◯◯◯ Very low	CRITICAL
Death from all causes (follow-up: range 48 weeks to 168 weeks; assessed with: incidence rate)											
9[1,2,3,4,5,6,7,8,9]	randomized trials	very serious[a]	very serious[b]	not serious	very serious[c]	none	22	2388	event rate 4.47 per 1000 person-years (1.42 to 7.91)	⨁◯◯◯ Very low	CRITICAL

Table S8 - Summary of findings (conclusion)

№ of studies	Certainty assessment						Effect			Certainty	Importance
	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	№ of events	№ of individuals	Rate (95% CrI)		
Death from ART-related adverse events (follow-up: range 48 weeks to 168 weeks; assessed with: incidence rate)											
9[1,2,3,4,5,6,7,8,9]	randomized trials	very serious[a]	very serious[b]	not serious	very serious[c]	none	4	2388	event rate 0.18 per 1000 person-years (0 to 1.56)	⨁◯◯◯ Very low	CRITICAL

Explanations

a. Eight studies with a high risk of bias and one with some concern about the risk of bias according to Cochrane's RoB2.

b. In the type of statistical analysis used in this review, very wide credibility intervals suggest high heterogeneity. Furthermore, there is heterogeneity due to clinical differences (mainly due to the presence of different ARV regimens, in which the same regimen is repeated in a maximum of two arms of the included studies), blinding and funding.

c. Credibility intervals are quite wide.

d. Four studies with a high risk of bias and one with some concern about the risk of bias according to Cochrane's RoB2.

e. Three studies with a high risk of bias and one with some concern about the risk of bias according to Cochrane's RoB2.

f. One study with a high risk of bias according to Cochrane's RoB2.

g. Two studies with a high risk of bias and one with some concern about the risk of bias according to Cochrane RoB2.

h. Three studies with a high risk of bias according to Cochrane's RoB2.

i. Four studies with a high risk of bias according to Cochrane's RoB2.

j. Seven studies with a high risk of bias and one with some concern about the risk of bias according to Cochrane's RoB2.

k. One study at high risk of bias and one with some concern about the risk of bias per Cochrane RoB2.

References

1.LOCKMAN, S. *et al*. Antiretroviral Therapies in Women after Single-Dose Nevirapine Exposure. New England Journal of Medicine; 2010.

2.LOCKMAN, S. *et al*. Nevirapine- versus lopinavir/ritonavir-based initial therapy for HIV-1 infection among women in Africa: a randomized trial. PLoS medicine; 2012.

3.RASHBAUM, B. *et al*. Darunavir/cobicistat/emtricitabine/tenofovir alafenamide in treatment-naïve patients with HIV-1: subgroup analyses of the phase 3 AMBER study. HIV Research & Clinical Practice; 2019.

4.KITYO, C. *et al*. Switching to Fixed-Dose Bictegravir, Emtricitabine, and Tenofovir Alafenamide (B/F/TAF) in Virologically Suppressed HIV-1 Infected Women: A Randomized, Open-Label, Multicenter, Active-Controlled, Phase 3, Noninferiority Trial. JAIDS Journal of Acquired Immune Deficiency Syndromes; 2019.

5.NAICKER, N. *et al*. Efficacy and Safety of Tenofovir-Containing Antiretroviral Therapy in Women who Acquired HIV while Enrolled in Tenofovir Gel Prophylaxis Trials. Antiviral Therapy; 2017.

6.HUHN, G. D. *et al*. Darunavir/cobicistat/emtricitabine/tenofovir alafenamide in treatment-experienced, virologically suppressed patients with HIV-1: subgroup analyses of the phase 3 EMERALD study. AIDS Research and Therapy; 2019.

7.IBRAHIM, F. *et al*. Bone mineral density, kidney function, weight gain and insulin resistance in women who switch from TDF/FTC/NNRTI to ABC/3TC/DTG. HIV Medicine; 2020.

8.ORRELL, C. *et al*. Fixed-dose combination dolutegravir, abacavir, and lamivudine versus ritonavir-boosted atazanavir plus tenofovir disoproxil fumarate and emtricitabine in previously untreated women with HIV-1 infection (ARIA): week 48 results from a randomised, open-label, non-inferiority, phase 3b study. The Lancet HIV; 2017.

9.SQUIRES, K. *et al*. Integrase inhibitor versus protease inhibitor based regimen for HIV-1 infected women (WAVES): a randomised, controlled, double-blind, phase 3 study. The Lancet HIV; 2016.

10.FIRNHABER, C. *et al*. Differences in antiretroviral safety and efficacy by sex in a multinational randomized clinical trial. HIV Clinical Trials; 2015.