Article details: 2014-0017	
Title	**The effects of HIV-1 subtype and ethnicity on CD4 decline in antiretroviral naïve patients: a Canadian-European collaborative cohort study**
Authors	Marina Klein, James Young, David Dunn, Bruno Ledergerber, Caroline Sabin, Alessandro Cozzi-Lepri, Francois Dabis, P Harrigan, Darryl Tan, S. Walmsley, M John Gill, Curtis Lindsey Cooper, Alexandra Scherrer, A. Mocroft, Robert Hogg, Fiona Smaill
Reviewer 1	**Marian Laderoute**
Institution	Immune system Management, Laboratory
General comments	While this paper addresses differences in CD4 decline by ethnicity accounting for clade of HIV-1 and includes Canadian data (about 16 %), it is of medical importance generally, because it identifies that there are unknown HOST factors (genetic, social and/or environmental) which impact immune function (and decline) of patients which clearly need to be identified. It appears to be an extension of earlier papers on the Swiss HIV Cohort Study (Muller V et al, AIDS 2009, May M et al, J. Inf.Diseases 2009) which also reported slower CD4 decline. What is new is, with a larger sample size and Canadian data, no evidence was found for informative censoring due to early ART initiation (potentially related to more females in the black population studied who generally have lower CD4 starting counts). For the field of HIV-1 treatment and prevention globally, while there is a move towards "disregarding the CD4 count" and initiating ART asap, in fact, having proper CD4 decline rates by treatment population is key to estimating the impact of the substantial cost of providing such a therapy to these populations (see Granich R et al, PLoS One, 2012, also benefits on TB risks see Pretorius C et al, AIDS 2014). As well these CD4 decline rates could be and have been used to estimate if immigrants became infected after immigration (Rice BD et al, AIDS 2012) important for other policy decisions. Other interesting and somewhat contrasting work in a similar area is the paper by McKinnon LR et al, PLoS One 2012) which shows Clade D (rarer) in Kenyans may be associated with enhanced rates of CD4 decline over A1 using similar methods (accounting for HLA protective alleles). While the analysis in this paper did not include Clade D, there were too few which WAS addressed in the limitations. This paper must be revised to include p values as these are advanced statistical methods, not available to reviewers or readers. There was one aspect that surprised me, which was in Table 6, the reference Canadian, male, non-black would have a HIV viral RNA load of 4 log 10 copies instead of 4.9 log 10 copies recently reported (see Cescon A et al, PLoS One 2013). This might bias or enhance the ability to show decreased CD4 decline in blacks (as people with VL of 10,000 are probably better off than those with 100,000). However it could be related to younger age in the current paper although in Table 3, blacks had 4.0 log copies per ml vs 4.4 in non-blacks, which probably was significant. How the reference viral load would impact the analysis should be addressed nonetheless. Marian Laderoute, PhD Medical Sciences-Immunology Immune System Management Clinic and Laboratory Ottawa No conflict of interest.
Reviewer 2	**Jody Boffa**
Institution	University of Calgary, Community Health Sciences
General comments	This paper is about the effects of black race and HIV-subtypes (circulating in Europe and Canada) on CD4 decline over time among ART-naïve people living with HIV (PLWH). The findings suggest that black patients exhibited slower CD4 decline compared to a combination of all other races. Researchers also reported a marginal effect of HIV sub-type on CD4 decline, but suggest a potential modifying effect of sub-type when combined with race. This manuscript is very well written with the use of appropriate modelling and a number of sensitivity tests to confirm the investigators' hypothesis. It also involves the combination of a number of European and Canadian cohorts (which would require an enormous amount of data collection and transformation), thereby strengthening its power to detect differences, and is the first of its kind to report on sub-type decline by race. Major comments: 1. We want to caution the investigators with regard to wording. This is not an appeal for political correctness, but rather a concern with unintentionally propagating stereotypes by use of convenient language. While we feel this paper contributes quality information to the HIV literature, there is an issue with using the term "ethnicity" in this paper. In the sociologic world, "black" is not generally accepted as an ethnicity. Ethnicity is defined as a similar culture, language, or religion, which many would argue is not static among the world's black populations. Whether or not one agrees with the existence or categorisation of "race," this paper appears to be defining people by

qualities associated with the concept of race, and not necessarily by epidemiology. May we suggest the re-categorisation of data by WHO regions (see: http://www.who.int/about/regions/en/)? This might better reflect epidemiologic similarities and may better explain CD4 decline. The difference reported among Caribbean vs African populations after sensitivity testing is a good example of how regional differences affect epidemiology (although we must also note that we were not persuaded by the assumption that HIV+ Carib peoples are predominantly black [ref Pg 7, line 28-30] – is there any reference in the literature to HIV status of Carib peoples in Canada by race? Replacing race with WHO region could eliminate this criticism). For the benefit of readers, it might also be important to explain up front why region of origin would be more epidemiologically relevant (i.e. not just proximity, but similar diet, socio-economic status (SES), cultures, prevalence of HIV, modes of transmission, co-infections such as parasites, etc).

2. Although the authors acknowledge the lack of SES as a covariate, (pg 15, lines 32-37), we would like to suggest the inclusion of a proxy variable in the analysis so as to strengthen (or disprove) the hypothesis. We remain cautious of using race as a proxy, as many readers may not have the time to consider the interplay between race and SES in their own interpretation of the paper (and more explicitly if using just the title or abstract as a reference), and we therefore feel it is important for the authors to do the heavy lifting on behalf of the reader. Many theorists discuss how poverty increases the risk of HIV transmission, leading to a higher prevalence of HIV and to heterosexual transmission (both resulting in greater risk of exposure) in Africa, regardless of sexual beliefs and practices. We understand that SES can be very difficult to capture in a retrospective study, yet it is important to recognize that SES could explain slower decline in CD4 count better than race. We suggest that the investigators conduct a sensitivity analysis by post code (where available) or include another SES proxy as a covariate. Other examples may include one or a combination of the following: income-level, occupation, education, sex, IDU, World Bank status of country of origin http://www.gfmag.com/tools/global-database/economic-data/12066-countries-by-income-group.html#axzz2wxu1JAmO - perhaps not the best proxy for well-educated immigrants – or even area-based proxies such as average persons per room, Townshend index, literacy levels, or population growth. The inclusion of SES could also dispel critiques pertaining to the healthy migrant effect (mentioned on page 15, line 44-47). In the absence of any available SES indicators (although we note that a combination of sex and IDU might be possible at minimum, given the reported data), we kindly ask the authors to be more cautious about attributing findings to region of origin. Perhaps this can be achieved by acknowledging the important contribution of SES in the introduction and throughout the paper, or even by declaring "black" as a proxy for SES from the start, with appropriate references from sociology.

3. With reference to the issue of power to detect the presence of modification (page 11 line 29,31), may we suggest (if the data make it possible) the use of the egen (group) command in stata for generating a smaller list of realistic combinations of variables in order to assess some degree of effect modification in a single model? This command can help identify the most common combinations amongst exposures and covariates, rather than all possible combinations, many of which may not actually exist in the dataset. Groupings then make it possible to include fewer variables in the model, while still examining possible effects of modification, e.g. common combinations might be African*SubTypeA*Hetero, Other*SubTypeB*MSM, etc.

Minor comments

1. Methods: page 7, lines 23,25: Could it be that the findings are confounded by the fact that the investigators had to lump all "other ethnicities" together? While we recognise that the combination of many different retrospective datasets can make these analyses near to impossible, we cannot be sure of the contributions of other region/ethnic groups if these data could be disaggregated. Were there enough patients to consider First Nations separate from "other ethnicities" amongst the Canadian data? Generally this group has a much different epidemiology than caucasian men who have sex with men (MSM), e.g.. A sensitivity analysis of this kind may also be a good proxy for lower SES among Canadian PLWH.

2. Interpretation: page 14, line 41-48, the chronic phase of infection corresponds to lower viral loads compared to the time of HIV acquisition and the syndrome phase, and thus we feel that the interpretation that a slower decline during the chronic phase "increases the opportunity for HIV transmission" is unsubstantiated.

	3. Table 6: Is it possible that starting at a lower CD4 count explains slower rate of decline (ie non-ARV interventions such as co-trimoxazole, nutritional interventions, etc, have a more rapid effect on persons with lower CD4 counts compared to persons starting at higher counts). We are not clear on how to test this from your data, but perhaps this could be an alternative explanation that is considered under limitations. (Does page 8 lines 27-29 refer to this? If so maybe a few more words on how this modelling was performed and how it would control for this issue.) 4. Table 3: We suggest using the more inclusive term "MSM" (men who have sex with men) over homosexual (it is generally recognised in the literature that some men who have sex with men - and also women - identify as heterosexual) Jointly reviewed by: Ms. Jody Boffa is a PhD candidate in Epidemiology specializing in TB and HIV at the University of Calgary and lecturer in Qualitative Methods and Epidemiology in the Nelson Mandela School of Medicine at the University of KwaZulu-Natal in Durban, South Africa. She has no competing interests to declare. Dr. Tsholofelo Mhlaba is a Public Health Physician, PhD student, and Lecturer in Health Service Research and undergraduate medical education at the Nelson Mandela School of Medicine at the University of KwaZulu-Natal in Durban, South Africa. She has no competing interests to declare. (mhlaba@ukzn.ac.za)
Author response	Response to Reviews **Reviewer 1: Marian Laderoute, PhD** Medical Sciences-Immunology Immune System Management Clinic and Laboratory Ottawa. No competing interests declared. **Comments to the Author** While this paper addresses differences in CD4 decline by ethnicity accounting for clade of HIV-1 and includes Canadian data (about 16 %), it is of medical importance generally, because it identifies that there are unknown HOST factors (genetic, social and/or environmental) which impact immune function (and decline) of patients which clearly need to be identified. It appears to be an extension of earlier papers on the Swiss HIV Cohort Study (Muller V et al, AIDS 2009, May M et al, J. Inf.Diseases 2009) which also reported slower CD4 decline. What is new is, with a larger sample size and Canadian data, no evidence was found for informative censoring due to early ART initiation (potentially related to more females in the black population studied who generally have lower CD4 starting counts). For the field of HIV-1 treatment and prevention globally, while there is a move towards "disregarding the CD4 count" and initiating ART asap, in fact, having proper CD4 decline rates by treatment population is key to estimating the impact of the substantial cost of providing such a therapy to these populations (see Granich R et al, PLoS One, 2012, also benefits on TB risks see Pretorius C et al, AIDS 2014). As well these CD4 decline rates could be and have been used to estimate if immigrants became infected after immigration (Rice BD et al, AIDS 2012) important for other policy decisions. Other interesting and somewhat contrasting work in a similar area is the paper by McKinnon LR et al, PLoS One 2012) which shows Clade D (rarer) in Kenyans may be associated with enhanced rates of CD4 decline over A1 using similar methods (accounting for HLA protective alleles). While the analysis in this paper did not include Clade D, there were too few which WAS addressed in the limitations. 1. This paper must be revised to include p values as these are advanced statistical methods, not available to reviewers or readers. [Editor's note: where 95% CI are presented, no p values are required.] **We provide confidence intervals for all inferential statistics.** 2. There was one aspect that surprised me, which was in Table 6, the reference Canadian, male, non-black would have a HIV viral RNA load of 4 log 10 copies instead of 4.9 log 10 copies recently reported (see Cescon A et al, PLoS One 2013). This might bias or enhance the ability to show decreased CD4 decline in blacks (as people with VL of 10,000 are probably better off than those with 100,000). However it could be related to younger age in the current paper although in Table 3, blacks had 4.0 log copies per ml vs 4.4 in non-blacks, which probably was significant. How the reference viral load would impact the analysis should be addressed nonetheless. The reference patient was created by centering covariates around a suitable value.

	Centering covariates tends to reduce the correlation between covariates and this can speed up convergence when fitting complex models and may improve the precision of estimates. So the reference value of 4 log 10 copies was chosen arbitrarily, although it is close to the median first HIV RNA measurement in many cohorts. Centering has no real disadvantages (see Kraemer & Blasey 2004, International Journal of Methods in Psychiatric Research, 13:141-151). **We probably can't include an additional table with patient characteristics cross classified by cohort but for the Canadian cohorts, the median first HIV RNA measurement was (cohort, log 10 copies): Calgary, 3.5; Hamilton, 3.7; Montreal, 4.0; Ottawa, 3.9; Toronto 4.4; Vancouver 4.7.** Reviewer 2: Joint review Ms. Jody Boffa is a PhD candidate in Epidemiology specializing in TB and HIV at the University of Calgary and lecturer in Qualitative Methods and Epidemiology in the Nelson Mandela School of Medicine at the University of KwaZulu-Natal in Durban, South Africa. She has no competing interests to declare. Dr. Tsholofelo Mhlaba is a Public Health Physician, PhD student, and Lecturer in Health Service Research and undergraduate medical education at the Nelson Mandela School of Medicine at the University of KwaZulu-Natal in Durban, South Africa. She has no competing interests to declare. (mhlaba@ukzn.ac.za) **Comments to the Author** This paper is about the effects of black race and HIV-subtypes (circulating in Europe and Canada) on CD4 decline over time among ART-naïve people living with HIV (PLWH). The findings suggest that black patients exhibited slower CD4 decline compared to a combination of all other races. Researchers also reported a marginal effect of HIV sub-type on CD4 decline, but suggest a potential modifying effect of sub-type when combined with race. This manuscript is very well written with the use of appropriate modelling and a number of sensitivity tests to confirm the investigators' hypothesis. It also involves the combination of a number of European and Canadian cohorts (which would require an enormous amount of data collection and transformation), thereby strengthening its power to detect differences, and is the first of its kind to report on sub-type decline by race. Major comments: 1. We want to caution the investigators with regard to wording. This is not an appeal for political correctness, but rather a concern with unintentionally propagating stereotypes by use of convenient language. While we feel this paper contributes quality information to the HIV literature, there is an issue with using the term "ethnicity" in this paper. In the sociologic world, "black" is not generally accepted as an ethnicity. Ethnicity is defined as a similar culture, language, or religion, which many would argue is not static among the world's black populations. Whether or not one agrees with the existence or categorisation of "race," this paper appears to be defining people by qualities associated with the concept of race, and not necessarily by epidemiology. May we suggest the re-categorisation of data by WHO regions (see: http://www.who.int/about/regions/en/)? This might better reflect epidemiologic similarities and may better explain CD4 decline. The difference reported among Caribbean vs African populations after sensitivity testing is a good example of how regional differences affect epidemiology (although we must also note that we were not persuaded by the assumption that HIV+ Carib peoples are predominantly black [ref Pg 7, line 28-30] – is there any reference in the literature to HIV status of Carib peoples in Canada by race? Replacing race with WHO region could eliminate this criticism). For the benefit of readers, it might also be important to explain up front why region of origin would be more epidemiologically relevant (i.e. not just proximity, but similar diet, socio-economic status (SES), cultures, prevalence of HIV, modes of transmission, co-infections such as parasites, etc). [Editor's note: please see Editors' comments (above) re: terminology and SES.] Please see response to Editors comments 2 and 3 above. 2. Although the authors acknowledge the lack of SES as a covariate, (pg 15, lines 32-37),

	we would like to suggest the inclusion of a proxy variable in the analysis so as to strengthen (or disprove) the hypothesis. We remain cautious of using race as a proxy, as many readers may not have the time to consider the interplay between race and SES in their own interpretation of the paper (and more explicitly if using just the title or abstract as a reference), and we therefore feel it is important for the authors to do the heavy lifting on behalf of the reader. Many theorists discuss how poverty increases the risk of HIV transmission, leading to a higher prevalence of HIV and to heterosexual transmission (both resulting in greater risk of exposure) in Africa, regardless of sexual beliefs and practices. We understand that SES can be very difficult to capture in a retrospective study, yet it is important to recognize that SES could explain slower decline in CD4 count better than race. We suggest that the investigators conduct a sensitivity analysis by post code (where available) or include another SES proxy as a covariate. Other examples may include one or a combination of the following: income-level, occupation, education, sex, IDU, World Bank status of country of origin http://www.gfmag.com/tools/global-database/economic-data/12066-countries-by-income-group.html#axzz2wxu1JAmO - perhaps not the best proxy for well-educated immigrants – or even area-based proxies such as average persons per room, Townshend index, literacy levels, or population growth. The inclusion of SES could also dispel critiques pertaining to the healthy migrant effect (mentioned on page 15, line 44-47). In the absence of any available SES indicators (although we note that a combination of sex and IDU might be possible at minimum, given the reported data), we kindly ask the authors to be more cautious about attributing findings to region of origin. Perhaps this can be achieved by acknowledging the important contribution of SES in the introduction and throughout the paper, or even by declaring "black" as a proxy for SES from the start, with appropriate references from sociology. [Editor's note: please see Editors' comments (above) re: SES and terminology] Please see response to Editors comment 8 above. Because we were unable to carry out suitable covariate adjustment for socio-economic status, we have been appropriately cautious when discussing ethnic differences. Like the reviewers, we suspect ethnicity is in part a marker for sociologic factors. In the Interpretation, we write: "Yet it remains unclear how ethnicity could affect the underlying pathogenesis of HIV infection" and "It is also possible that ethnicity simply serves as a marker for socioeconomic, cultural and environmental factors that may influence immunologic progression" and we conclude the paper by stating that "While we have some evidence of differences in CD4 decline between subtypes, particularly in patients of African ancestry, research to uncover the underlying biologic or sociologic reasons for slower immunologic progression among patients of African ancestry is warranted." **That is, we think the reason for ethnic differences is still an open question and this view is clearly stated in the final paragraph.** 3. With reference to the issue of power to detect the presence of modification (page 11 line 29,31), may we suggest (if the data make it possible) the use of the egen (group) command in stata for generating a smaller list of realistic combinations of variables in order to assess some degree of effect modification in a single model? This command can help identify the most common combinations amongst exposures and covariates, rather than all possible combinations, many of which may not actually exist in the dataset. Groupings then make it possible to include fewer variables in the model, while still examining possible effects of modification, e.g. common combinations might be African*SubTypeA*Hetero, Other*SubTypeB*MSM, etc. **Our strategy for selecting covariates was to use those seen to be relevant in a recent study (ref [23], Touloumi et al, Clinical Infectious Diseases 2013), coding each covariate in the same way, so that readers can compare results between studies.** Minor comments 1. Methods: page 7, lines 23,25: Could it be that the findings are confounded by the fact that the investigators had to lump all "other ethnicities" together? While we recognise that the combination of many different retrospective datasets can make these analyses near to impossible, we cannot be sure of the contributions of other region/ethnic

	groups if these data could be disaggregated. Were there enough patients to consider First Nations separate from "other ethnicities" amongst the Canadian data? Generally this group has a much different epidemiology than caucasian men who have sex with men (MSM), e.g.. A sensitivity analysis of this kind may also be a good proxy for lower SES among Canadian PLWH. **We are reluctant to carry out exploratory analyses of small subsets of our data. We report only one exploratory analysis – Figure 3 (previously Figure 1b) – and this is of the subtype B which has many more patients than other subtypes. Here we cross-classify patients with subtype B by ethnicity and region giving three categories: black patients from the Caribbean, black patients from other regions (mostly Africa), and patients of other ethnicities. Even then, one of these categories contains only 73 patients. This is the problem with such exploratory analyses – the subgroups end up so small that their statistics may not be reliable.** 2. Interpretation: page 14, line 41-48, the chronic phase of infection corresponds to lower viral loads compared to the time of HIV acquisition and the syndrome phase, and thus we feel that the interpretation that a slower decline during the chronic phase "increases the opportunity for HIV transmission" is unsubstantiated. **While higher viral loads seen in acute infection are clearly associated with increased risk of transmission per episode of contact, there is considerable debate about what pool of HIV infected persons is responsible for most of HIV transmission events (e.g. what are the relative contributions of high viremia for a short duration vs. lower but still considerable viremia over longer durations). Different studies have arrived at widely differing estimates of the proportion of HIV transmission events that occur during acute HIV infection, ranging from 5% to 95%. In settings where the predominate mode of HIV transmission is heterosexual and serial monogamy is more common in sexual partnerships, the opportunity for HIV transmission will be increased for individuals who are infected, viremic and untreated over long periods of time (Cohen, Plos Medicine 2011). In other words, the asymptomatic stage of infection will typically contribute more to the net transmission of HIV-1 over the lifetime of an infected individual, because of its longer duration (Hollingsworth JID 2008). Furthermore, our statement is a relative one: those with slower CD4 decline (African ancestry) have increased opportunity for HIV transmission compared to those with faster decline (who become symptomatic and/or are offered treatment sooner). Even models that show acute HIV infection is a major contributor to transmissions (Powers et al Lancet 2011) estimate that 38.5% of transmissions are accounted for by acute infection (meaning the majority, 61.5% are due to chronic infections.) So while early infection is critical to widespread HIV infection in Africa, chronic infection still likely explains a greater number of transmissions overall. We have added references to support this.** 3. Table 6: Is it possible that starting at a lower CD4 count explains slower rate of decline (ie non-ARV interventions such as co-trimoxazole, nutritional interventions, etc, have a more rapid effect on persons with lower CD4 counts compared to persons starting at higher counts). We are not clear on how to test this from your data, but perhaps this could be an alternative explanation that is considered under limitations. (Does page 8 lines 27-29 refer to this? If so maybe a few more words on how this modelling was performed and how it would control for this issue.) **It's not entirely clear what the reviewers are referring to. Models do assume a linear rate of decline in square root CD4 cell count during the stable chronic phase of HIV infection and this approach is consistent with the vast majority previous studies of CD4 decline. We did not evaluate the use of other non-ARV interventions in this study but given that starting CD4 cell counts for all ethnicities were well above the threshold for which these interventions are indicated, it's hard to see how their use would have influenced our results. Our selection criteria did remove measurements that were likely to be made in the advanced stages of HIV or when there were likely to be clinical symptoms. The section describing this process has been amended as per Editor's comment 6, which hopefully should clarify this further.**

	4. Table 3: We suggest using the more inclusive term "MSM" (men who have sex with men) over homosexual (it is generally recognised in the literature that some men who have sex with men - and also women - identify as heterosexual) **This has been done as indicated above.**