Dear Sir

We appreciated the letter sent by Ciccullo and Colleagues[1] in response to our paper recently published on your Journal[2]. They shared the data from their cohort of DTG-treated PLWH, carrying no evidence for a difference in the risk of DTG discontinuation due to toxicity according to sex.

Reproducibility in epidemiological studies is of great importance: data confirmed by different settings are undoubtfully more reliable that single cohort data. Nevertheless, we believe that comparison between results of different studies without having access to the data from all of them is always difficult because can be due to a large number of factors including the existence of effect modifiers, use of not standardized data collection and definition of outcomes as well as different ways of dealing with measured confounding and potential existence of other hidden bias (selection/collider).

We believe that our analysis of ICONA cohort data, which focused specifically on the association between sex and response to DTG- containing regimens, cannot be directly compared to those of Ciccullo study.

First of all, the inclusion criteria and study population. We designed different analyses for the ART-naïve and ART-experienced population while in the Odoacre cohort the analysis includes a mixed population of ART-naïve and ART-experienced patients. Although our estimates of the RH were similar in the two analyses, it seems methodologically unsound to merge together populations which have very different therapeutic history and in whom different set of predictors may act as potential confounders.

Second, for the classification of discontinuations, we used the primary reason reported by the treating physicians, as coded in the ICONA database that follows the international HICDEP standard for data collection on HIV observational studies (https://www.hicdep.org/), distinguishing failure, from toxicity, simplification and other reasons (pregnancy, planned pregnancy, inclusion in trial, unspecified, other). We do not know how reasons of discontinuation are collected by the Ciccullo cohort; in particular, it is unclear whether discontinuations due to simplification were counted as events in the analysis with outcome treatment failure.

According to our classification, in our paper we analyzed treatment discontinuations for toxicity as classified by clinicians by three definitions: i) including all causes of toxicity or intolerance; ii) by considering only neuropsychiatric (NPS) toxicity; and, iii) to be more conservative, discontinuation for any cause apart treatment simplification. In experienced PLWH we also had the power to disentangle discontinuations according to the DTG regimen (2DR vs 3DR). All analyses which have not been replicated in the Ciccullo study.

Both in ART-naïve and -experienced settings of our study, women had a higher risk of discontinuing DTG for toxicity and for all except simplification reasons, after adjustment for age and nation at birth. Further, in experienced PLWH, women had a higher risk of discontinuation, particularly in case of dual-DTG regimens. Nevertheless, the rate of discontinuation was so low, less than 9% by 4 years, that the risk difference was negligeable. Moreover, in our study we presented some important outcomes of pregnancy during DTG that would be also interesting to replicate in Odoacre .

One key limitation in the analysis of data of observational studies is bias due to confounding. Although sex is established at birth and therefore potentially unconfounded, our analysis was further controlled for age and nation at birth (variables known to be associated with the risk of ART discontinuation) and we even described how the results were robust against potential source of unmeasured confounding. Again, in the analysis of Ciccullo et al, it is not clear how potential bias due to measured confounding was minimized.

Finally, difference in the results of association analyses may arise by the presence of effect measure modifiers. We compared the main characteristics of our ART-naïve cohort with that of the Odoacre study. ART-naive women included in our study were older and with longer history of HIV. The average age of PLWH included in Odoacre was 10 years older than that of our participants (50 vs. 40 years old) and age is a known factor affecting adherence to ART. Also women were significant older than men in our study and we do not know the age distribution by sex from the Table provided by the Odoacre authors. Nadir CD4 count was also significantly lower in the Odoacre analysis compared to our participants (200 vs. 341 cells/mm^3). Again, CD4 count nadir is a known predictor of ART discontinuation and might have acted as an important effect measure modifier for the association with sex.

In conclusion, we appreciate the work by Ciccullo et al to reproduce our results but without performing the exact same analysis in both datasets, reasons for discrepancies between the two studies could be due to a large number of factors and possible explanations can only be of

speculative nature and do not really help us understanding whether there is a genuine higher risk of discontinuation of DTG due to toxicity in women vs men.

References

1- Arturo CICCULLO[1]*, Gianmaria BALDIN[2], Gaetana STERRANTINO[3], Giordano MADEDDU[4], Gabriella D'ETTORRE[5], Cristina MUSSINI[6], Simona DI GIAMBENEDETTO **Evaluating tolerability of dolutegravir-based regimens in a large clinical-practice cohort**

2- D'arminio Monforte A, Tavelli A, Sala M et al. Long-term outcome of dolutegravir-containing regimens according to sex: data from the ICONA study. J Antimicrob Chemother. 2023 Feb 13:dkad026.