First, thank you for your interest in our paper. We are glad you made those comments, both the general and precise ones. With this answer, we hope we will be able to further complete the discussion about selection effects and their extent in our sample.

The main point that we want to bring to the attention of the referee is that a full SNANA analysis in collaboration with Daniel Scolnic with the proposed results of Nicolas 2020 has started, and should be a whole paper in itself. We are eager to work with such a powerful simulation tool to have an even better insight concerning those effects and how they affect cosmology.

Concerning the precise points brought up to our attention, here are some elements of answers:

1 Of the surveys' spectroscopic follow-up

1. SNLS's detection efficiency $\varepsilon \approx 0$ for $i \approx 24.8\,\mathrm{mag}$

   A limiting magnitude of $m_{\mathrm{lim}} = 23.5\,\mathrm{mag} \Rightarrow z_{\mathrm{lim}} = 0.36$, which would lead to only 26/236 SNe instead of 102/236 with our current cut

2. HST may have a follow-up efficiency that we should take into account like we did for SDSS;

3. Misunderstanding about the 20% of SNf's SNe that had selection effects:

   The 80% of SNf's SNe that had no selection effects are the 114 SNe that are in our sample.

2 Of the x_1 bias that doesn't appear on m

Then, the referee insists on the fact that "Biases in x1 are expected as a function of redshift simply from survey modeling/selection effects", showing the following figure from Kessler & Scolnic 2017 to point out that biases in x_1 become apparent much sooner that biases on m.

Yet, from the same figure, biases in c appear at $z = 0$. Here we want to convince the referee that if we don't find any sign of color bias in our sample, we may consider little to no bias on x_1. We thus studied the x_1 and c distributions of the end of SDSS and the start of PS1, for $0.10 < z < 0.20$. In this redshift range, the SDSS cut dataset contains the most questionable SNe Ia, for the SNe between $0.15 < z < 0.20$ are between our conservative and fiducial cuts, due to limited spectroscopic resources; the PS1 dataset is however quite robust for these SNe are far from both the conservative and fiducial cuts; see Fig. 2.

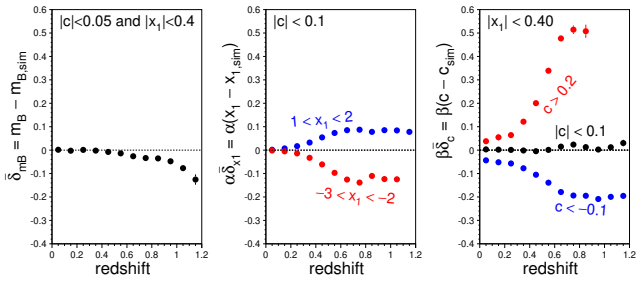

Fig. 1: Bias corrections $\bar{\delta}_{m_B}$, $\alpha\bar{\delta}_{x_1}$, and $\beta\bar{\delta}_c$ are shown as a function of redshift. The pre-factors α, β are used to show the bias in distance-modulus magnitudes. The parameter selection ranges are shown on each panel.

The results are shown Fig. 3. We represented the normed histograms along with their error bars, and found that they don't differ much. To ascertain this idea, we used a Kolmogorov-Smirnov test that gave us a p-value of 0.265 for the color, and 0.137 for the stretch; while it doesn't tell us that both the SDSS and PS1 parts come from the same distribution, as would be expected if we were free from selection effects (so that the samples were random draws from what Nature could give us), it is not excluded. This particuliar figure will appear upon re-submission of the paper for it adresses pertinent questions.

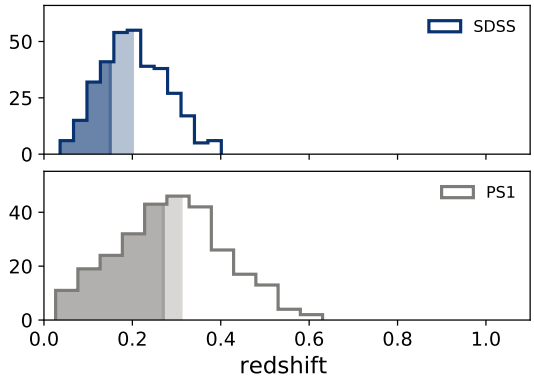

Fig. 2: Redshift histograms of SNe Ia from the SDSS and PS1 datasets respectively

Moreover, we tried to see if a particular trend of color evolution was visible, reproducting the Fig. 3 from the original N+20 paper plotting the color instead of the stretch without any modeling (Fig. 4, right), and recalled what it looked like for the stretch (Fig. 4, left). For additional information, we represented:

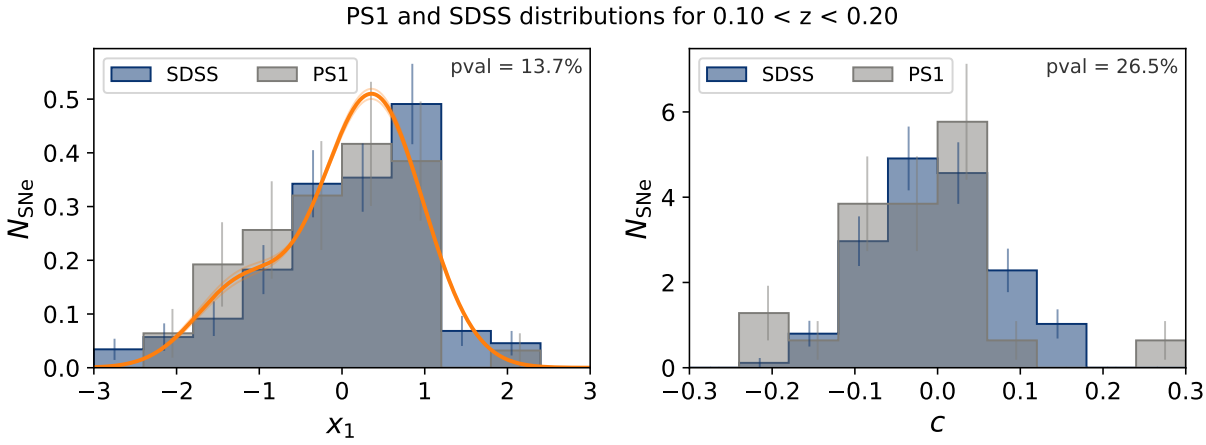

Fig. 3: x_1 and c normed histograms of the SDSS and PS1 samples for $0.10 < z < 0.20$ and their errorbars. On the left, we find in orange lines the Base model distributions for $z = 0.10, 0.15, 0.20$ from top to bottom. The p-values are the results from a Kolmogorov-Smirnov test; they don't show any indication that the samples are not taken from the same distribution, for both c and x_1.

- The full, uncut dataset in dark grey;
- The fiducial sample in light grey;
- The SNe above the fiducial cut in green (the SNe that we removed from the full sample to obtain the fiducial ones);
- The conservative sample il transparent light grey;
- And the fiducial sample without HST data.

The last one serves to show that even without HST data, the trend of the color to go up in the last bin is not governed by HST only. We find that the averages are compatible with a constant mean color, and that the fiducial and conservative samples are quite close in terms of mean values while removing the low-color trend that is expected from selection effects. Concerning the stretch, the fact that the full, fiducial and conservative samples are close are an indicator that the selection effect have maginal impact on mean x_1 in comparison to the drift.

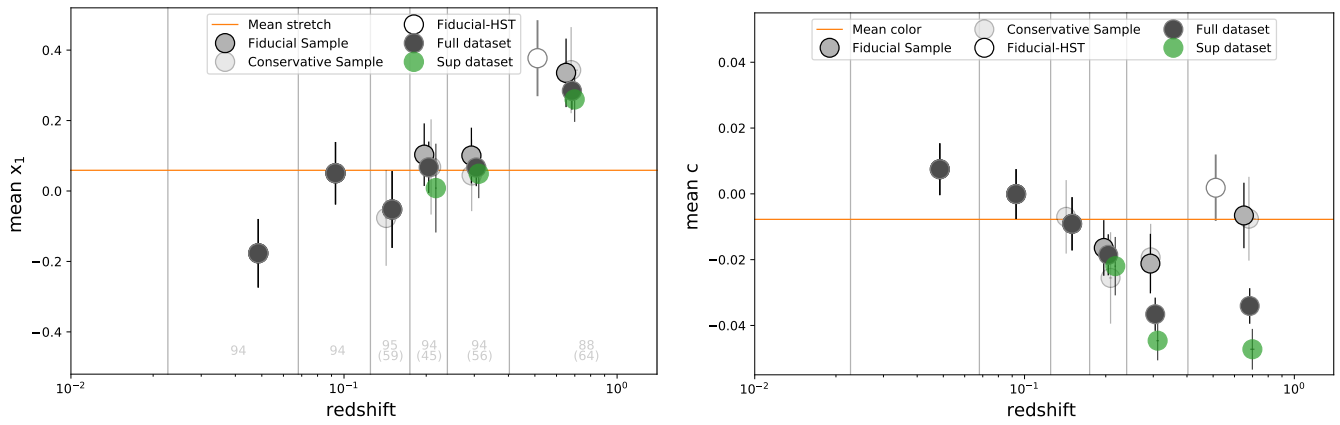

Fig. 4: Mean stretch and color of the complete sample in the same 6 bins of equal sample size (based on the fiducial sample). The dark grey markers represent the full, uncut sample; the light grey ones the fiducial sample; the SNe above the fiducial cut are in green; in transparent light grey is the conservative sample and the open marker shows the means for the complete sample from which we removed HST data. For the color, we might think that there is an evolution for the 5 first bins, but the 6th one breaks it, even without HST data, while for the stretch, with our without HST data we see the same evolution

3 Of the x_1 drift of simulated surveys

We recall here that a full SNANA analysis will soon be ongoing and should be the subject of a dedicated paper, giving more indepth insight concerning the simulations thoughts that were shared with us.

The referee pointed out that simulated samples without drift parameters do recover a stretch evolution, saying that the x_1 drift could be entirely explained by selection effects. We agree that both our supposed drift and selection effects go the same way, so it seems natural to see this trend in other studies. However, concerning the Fig. 6 of Mosher et al. 2014 which uses SDSS data we see that this trend actually goes up, according to selection effects, only after $z > 0.2$, which corresponds to our redshift cut for that survey. We feel that the plots the referee showed us are mostly examples that simulated samples include both drift and selection effects, the two of them stacking at the end of the redshift range.

Concerning the inclusion of the Per sample Asym model without using simulations beforehand, we want to point out that the idea is actually to compare models prescribing a drift parameter with those that don't, using pertinent models (Howell, Per sample Asym for instance). The Per sample Asym model is used in current cosmological studies and was said by their authors to be able to account for unmodeled redshift drifts thanks to the limited span of the surveys: we also wanted to test this affirmation. It would seem that this model is relevant for this study.

4 General comments

The parameters comparison in the referee's comments are indeed not agreeing with each other. There may be two reasons for that:

1. Even if we work on the same surveys, the datasets we use are really different from Scolnic & Kessler 2016 because they worked on their corrected full sample while we used a subset of that through the redshift cuts; differences in the computed parameters are expected in that matter.

2. Moreover, the parameters we say are in really good agreement are those of PS1 from Scolnic & Kessler 2018, in which they presented updated results that are in much better agreement with ours; we suspect that the same work done for SNLS and SDSS would go the same way.