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

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.

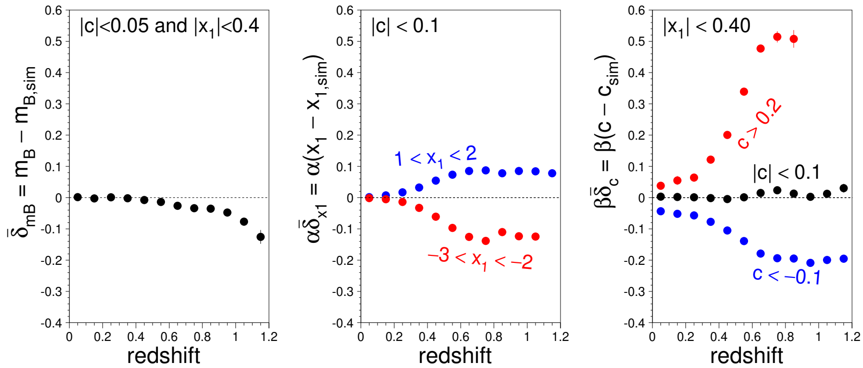

Fig. 1: Bias corrections $\bar{\delta}_{m_B}$, $\alpha\bar{\delta}_{x_1}$, and $\beta\bar{\delta}_c$ are shown as a function of redshift. The pre-factors α, β are used to show the bias in distance-modulus magnitudes. The parameter selection ranges are shown on each panel.

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.

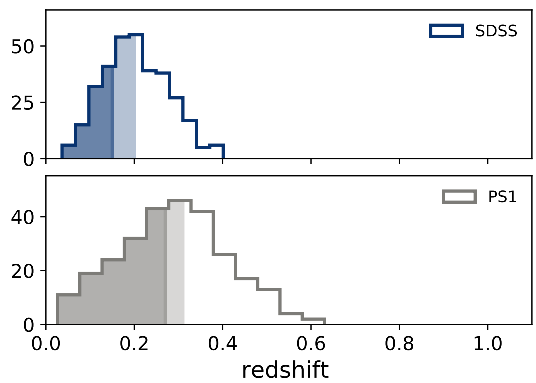

Fig. 2: Redshift histograms of SNe Ia from the SDSS and PS1 datasets respectively

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.

Moreover, we tried to see if a particular trend of color evolution was visible, reproducting the Fig. 3 from the original N+20 paper. We find that the averages are compatible with a constant mean color, see Fig. 4

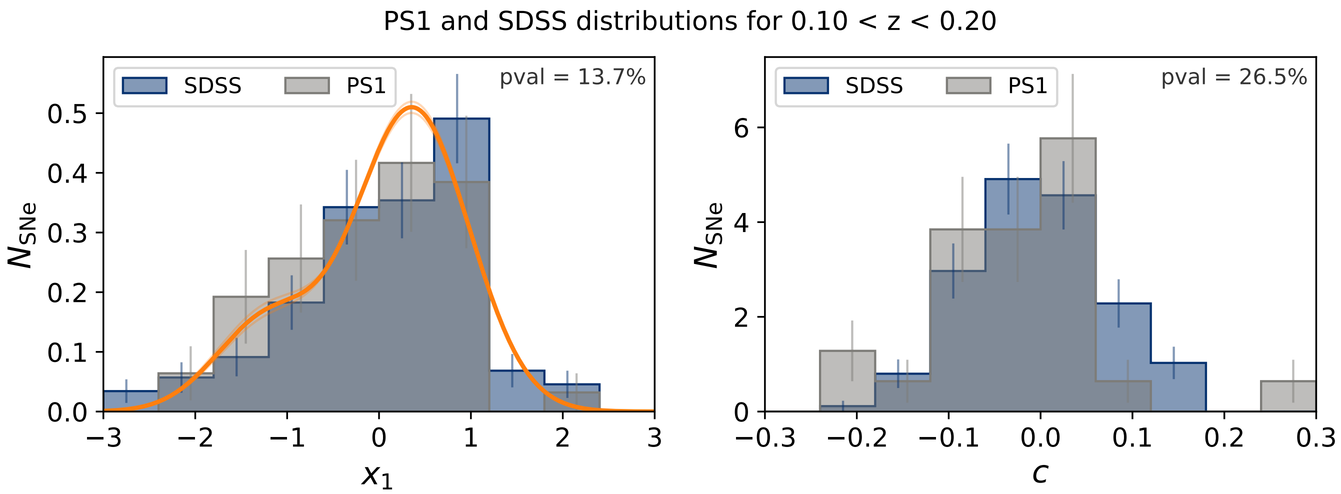

Fig. 3: x_1 and c normed histograms of the SDSS and PS1 samples for $0.10 < z < 0.20$ and their errorbars. On the left, we find in orange lines the Base model distributions for $z = 0.10, 0.15, 0.20$ from top to bottom. The p-values are the results from a Kolmogorov-Smirnov test; they don't show any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

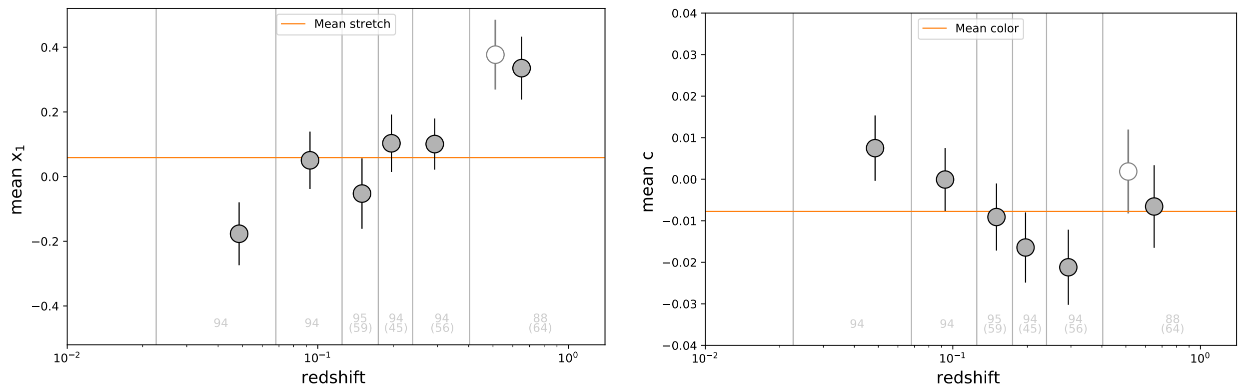

Fig. 4: Mean stretch and color of the complete sample in the same 6 bins of equal sample size. The open marker show the means for the complete sample from which we removed HST data. For the color, we might think that there is an evolution for the 5 first bins, but the 6th one breaks it, even without HST data, while for the stretch, with our without HST data we see the same evolution