Summary: We thank the reviewers for constructive comments, including noting that the work is non-trivial and is a needed contribution. We (re-)emphasize that the key contribution of the present work is the new *finite sample* consistency guarantee for a *non-parametric* EM algorithm; the key technical advance is the novel use of Gateaux derivatives to replace inner products in classical parametric EM theory. **Experiments:** We conducted experiments to address comments about the more challenging MNAR setting and comparison against additional baselines: we conclude that the utility is not limited to MCAR. We mimic MNAR in the bladder tumor dataset by artificially masking counts with probability ϵ conditional on zero counts ($\epsilon = 0.25, 0.3, 0.35$) or non-zero counts ($\epsilon = 0.2$). Figure 1a) has results. An increase in the zero missingness probability of 1% leads to an approximately 0.28% increase in the final estimate of expected tumors over MCAR: this is still useful for moderate differences in conditional missingness probabilities. For baselines, we include last value carried forward (LVCF), median, and complete case analysis (CCA). At study end, MCAR overestimates the expected number of tumors relative to using the full dataset by 0.11 ; LVCF underestimates by 0.08 and median and CCA both underestimate by 1.36. CCA has high variance. LVCF does well estimating the bladder tumor mean function, but handles discontinuities poorly. Figure 1b) shows results on a step function: LVCF shows higher bias after a discontinuity than EM. CCA again displays high variance. **Theory:** We clarify the non-trivial nature of theoretical extensions under departure from MCAR, for which the present work paves the way (line 32).

Reviewer 1: Lack of uncertainty quantification: this is challenging due to lack of asymptotic normality of the distance between the estimator and true mean function. We took the common approach of proving consistency and rate of convergence [18,30], leaving test statistics to future work [2]. **Linking c to r intuitively:** as the lower bound on expected number of cigarettes over any interval increases, we become more robust to initialization. This happens as minimum smoking risk and/or minimum interval sizes grow. Verifying that $r \leq \frac{c}{4}$ holds in practice requires knowing the true mean function. We recommend (line 135) trying multiple initializations and examining likelihood. Existing EM approaches [3, 35] have similar limitations. **Assumption 7:** is satisfied if $N(\tau)$ (e.g., cigarettes over study) is uniformly bounded. **Lack of missingness tolerance check:** we have this. Our theory holds for $\epsilon < \frac{c}{3b+c}$ (line 219). We need scientific intuition about b and c (uniform upper and lower bound on mean function increments) to apply it.

Reviewer 2 Thank you for your positive comments. We added two requested experiments in Figure 1.

Reviewer 3: Complete case analysis baseline: this requires deleting a participant's data starting from their first missing observation. With 5% missingness and four EMAs per day this on average loses all information after day 5 of a 15 day study, which is inefficient. Figure 1 shows the high variance of this method. **Observed data likelihood baseline:** it may not have a unique maximizer. Consider only one participant with three intervals with the middle missing. The observed data likelihood does not have a unique maximizing mean function (middle increment could be any non-negative value). Adding more participants with no observation times aligning with those of the first participant would still lack a unique maximizer. **Should show MAR and/or MNAR theory:** While asymptotic EM MAR results exist, the finite sample case is unstudied. Current finite sample results assume MCAR [3,35]. Our finite sample MCAR results for panel count are useful for smaller mHealth datasets. In our setting, proving local uniform strong concavity is difficult under MAR. Eqn 13 of the appendix relies on a constant ϵ and linearity of expectation. **MAR and sometimes MNAR full data distribution is identified:** thank you for the helpful references, which we will cite. Their setting is different: they have multiple observations of the same random variable. In ours, when sampling interval sizes from an absolutely continuous distribution, with probability 1 no two interval sizes across participants will be equal. Every count observation in the study may come from a different random variable. Further the references focus primarily on identification, but estimation poses additional challenges. (Malinsky et al. 2020) addresses estimation, but again in the setting of iid samples of the same random vector.

Reviewer 4: Baselines added (Fig. 1). **Lack of sensitivity to missingness:** we have this in Figure 4 of the supplement. **Comparison to literature:** work is limited on whether participants underestimate/overestimate their smoking count EMAs over long intervals. However our psychology coauthors who specialize in smoking consider this finding plausible.

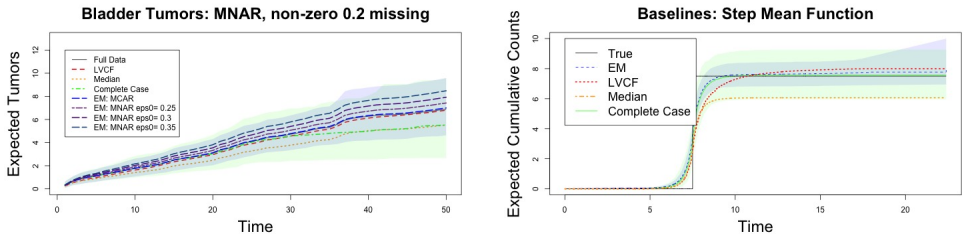

Figure 1: a) MNAR: Missingness ($\epsilon = 0.2$ for non-zero counts, varied ϵ for zero counts) results in minimal bias for our method. Other baselines treat the MCAR case. b) Step mean function: LVCF introduces bias compared to EM (blue).