We thank all the reviewers for their careful reading and thoughtful comments. We also thank the reviewers for general comments on the presentation, which we will address while preparing our final manuscript.

Reviewer 4 - Importance of removing $1/\varepsilon$ factor: We note that for algorithms with binary output (say YES or NO), larger values of ε provide rather vacuous guarantees. For example, $\varepsilon = 1.1$ would permit an algorithm to go from outputting YES with probability $1/4$ to outputting YES with probability $3/4$ by changing just a single point, which we do not consider private. Additionally, data analysis pipelines (e.g., model selection) in practice typically contain many private analyses, therefore, in order to achieve a reasonable overall privacy guarantee, each individual private algorithm must have a small ε as the privacy budget is split among the queries. For both these reasons, a minimal dependence on ε is preferred. We note that the $1/\varepsilon$ multiplicative baseline can be achieved with a simple application of the subsample and aggregate framework, a general purpose method for producing differentially private testing algorithms from non-private testing algorithms. The transition from a multiplicative ε factor to an additive ε factor has both theoretical and practical significance. In some regimes our private sample complexity is dominated by the non-private sample complexity (which never happens with a multiplicative dependence on ε). This implies much lower sample complexities in many important regimes, even for moderate sized ε: for example with $\varepsilon = 0.1$, subsample and aggregate requires 10x as much data as the non-private algorithm, while for some settings of d and α, our algorithm requires less than 2x as much data.

Reviewer 4 - Relevance to the community: We note that these problems are of core interest to the community, and most papers in this particular line on private hypothesis testing have appeared in either NeurIPS or ICML (see, e.g., Cai et al. ICML'17, Cummings et al. NeurIPS'18, Acharya et al. NeurIPS'18, Aliakbarpour et al. ICML'18, Aliakbarpour et al. NeurIPS'19).

Reviewer 6 - Practical importance: Our paper falls into the category of goodness-of-fit testing, which is ubiquitous in scientific research including studies that typically use sensitive information such as voting behaviour or clinical trials (e.g. [1, 4, 5]). The specific problems we study use the assumption that the analyst knows the family of distributions that the data come from (product or multivariate normal distributions). These types of parametric tests are often more powerful than non-parametric ones in the sense that they require fewer samples, and are thus often used in medical research [2]. In particular, testing the mean of a normal distribution is one of the most fundamental statistical primitives, most often achieved via a Z-test or T-test. (Two-tailed) Z-tests are mean tests for normal distributions with known covariance, which is exactly one of the problems we study in this paper, and they are standard tests used in studies of treatment effects [3, 6]. Since drug trials are paradigmatic of studies where the data contain highly sensitive information, this demonstrates the need for sample-efficient differentially private alternatives for this task. Finally, we would like to note that, based on our proof techniques, we generally expect that our algorithms would perform well in practice, even if the distribution is not exactly Gaussian, but rather "well-behaved" around the origin.

Reviewer 8 - Applicability of the approach: We believe that determining the optimal sample complexity is an important first step to the implementation of practical differentially private algorithms for these problems. Therefore, we consider our near-optimal with respect to sample complexity but computationally inefficient algorithm to be an important, non-trivial, first step towards this goal. Moreover, our computationally efficient algorithm is relatively simple to implement (as it consists of a truncation step, a filtering step, and then applying a variant of the popular chi-squared statistic). We predict that the performance of the algorithms in the wild will be faithful to their theoretical guarantees and close to their non-private counterparts. Empirical evaluation of private identity tests with similar dependence on ε in their theoretical guarantees shows that the cost of privacy is low in various settings (Cai et al. ICML'17).

References

[1] Aaron Blair, P Decoufle, and Dan Grauman. Causes of death among laundry and dry cleaning workers. *American journal of public health*, 69:508–11, 06 1979.

[2] Richard Chin and Bruce Y. Lee. Chapter 15 - Analysis of Data. In *Principles and Practice of Clinical Trial Medicine*, pages 325 – 359. Academic Press, New York, 2008.

[3] S.-C. Chow, J. Shao, H. Wang, and Y. Lokhnygina. Sample size calculations in clinical research. In *Statistical Theory and Related Fields*, Chapman & Hall/CRC Biostatistics Series. Taylor & Francis, 2018.

[4] J Gill, J Endres-Brooks, P Bauer, W.J. Marks, and Robert Montgomery. The effect of ABO blood group on the diagnosis of von willebrand disease. *Blood*, 69:1691–5, 07 1987.

[5] William A. Glaser. The family and voting turnout. *The Public Opinion Quarterly*, 23(4):563–570, 1959.

[6] K. K. Gordon Lan, Yuhwen Soo, Cynthia Siu, and Mey Wang. The use of weighted Z-tests in medical research. *Journal of Biopharmaceutical Statistics*, 15(4):625–639, 2005. PMID: 16022168.