Statistical design of quantitative mass spectrometry-based proteomic experiments

Ann L. Oberg*, Olga Vitek†‡

February 2, 2009

Supplementary materials

*Division of Biostatistics, Department of Health Sciences Research, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905; oberg.ann@mayo.edu

†Department of Statistics and Department of Computer Science, Purdue University, 250 N. University Street, West Lafayette, IN 47907; Olga Vitek, ovitek@stat.purdue.edu

‡ Corresponding author

Variances such as σ^2_{Indiv} and σ^2_{Error} discussed in the manuscript depend on numerous experimental protocols and settings, and can vary greatly between experiment types and labs. Here we investigate the effect of an increased experimental error on the performance of the designs. Specifically, we reproduce some of the plots from the main manuscript, while changing the values of $\sigma^2_{Error} = k \cdot \sigma^2_{Indiv}$, where k= 2, 3 and 5.

Section 3.2: Replication

The section considers the choice of biological versus technical replicates in a label-free workflow, with a completely randomized design. The variance of a comparison between groups is

$$Var(\bar{y}_H - \bar{y}_D) = 2\left(\frac{\sigma^2_{Indiv}}{I} + \frac{\sigma^2_{Prep}}{IJ} + \frac{\sigma^2_{Error}}{IJK}\right), \tag{1}$$

where I denotes the number of individuals in each group, J is the number of sample preparations per individual, and K is the number of runs per sample preparation.

Fig. 1 extends Fig. 6(b) in the main manuscript. We set σ^2_{Indiv} and σ^2_{Prep} to the median experimental values as in the mail manuscript, and illustrate the effect of increasing σ^2_{Error}. One can see that (1) an increase in σ^2_{Error} results in an overall increase in the variance of the comparison; (2) when σ^2_{Error}increases, additional technical replicates produce a stronger reduction of $Var(\bar{y}_H - \bar{y}_D)$, and (3) when the total number of runs is fixed, the advantage of biological replication holds. For example, in the panel (c) of the figure, an experiment without technical replicates which allocates 15 runs to 15 individuals in a group has a smaller variance than an experiment which allocates 15 runs to 5 individuals, 1 sample preparation and 3 technical replicates.

Section 3.3.1: Blocking in a label-free workflow

The section considers the effect of blocking in a label-free workflow. If we denote I the total number of individuals in a group, the variance of the comparison in a completely randomized design is

$$Var(\bar{y}_H - \bar{y}_D) = 2\left(\frac{\sigma^2_{Block} + \sigma^2_{Indiv} + \sigma^2_{Error}}{I}\right). \tag{2}$$

(a) (b) (c)

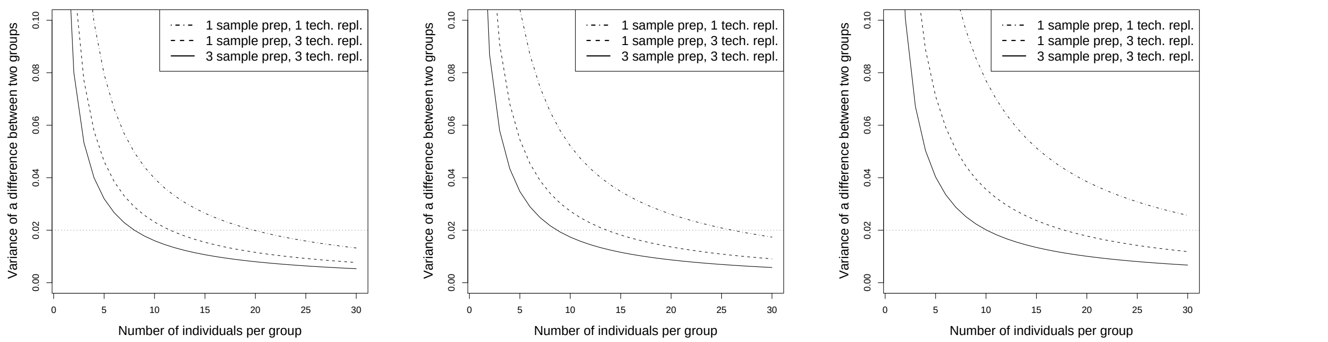

Figure 1: Replication in a label-free workflow: extends Fig. 6(b) in the main manuscript. (a) $\sigma^2_{Error} = 2\sigma^2_{Indiv}$; (b) $\sigma^2_{Error} = 3\sigma^2_{Indiv}$; (c) $\sigma^2_{Error} = 5\sigma^2_{Indiv}$.

and in the blocked design is

$$Var(\bar{y}_H - \bar{y}_D) = 2\left(\frac{\sigma^2_{Indiv} + \sigma^2_{Error}}{I}\right). \quad (3)$$

Both expressions depend on $\sigma^2_{Indiv}+\sigma^2_{Error}$, regardless of the relative magnitude of the two variances. Thus an increase in σ^2_{Error} will result in a systematic increase in $Var(\bar{y}_H - \bar{y}_D)$.

Fig. 2 extends Fig. 6(c) in the manuscript and illustrates this result. We set σ^2_{Indiv} and σ^2_{Block} to the same values as in Fig. 6(c), and increase σ^2_{Error}. One can see that this indeed results in the overall increase of the variances, but the relative performance of the designs is unchanged.

(a) (b) (c)

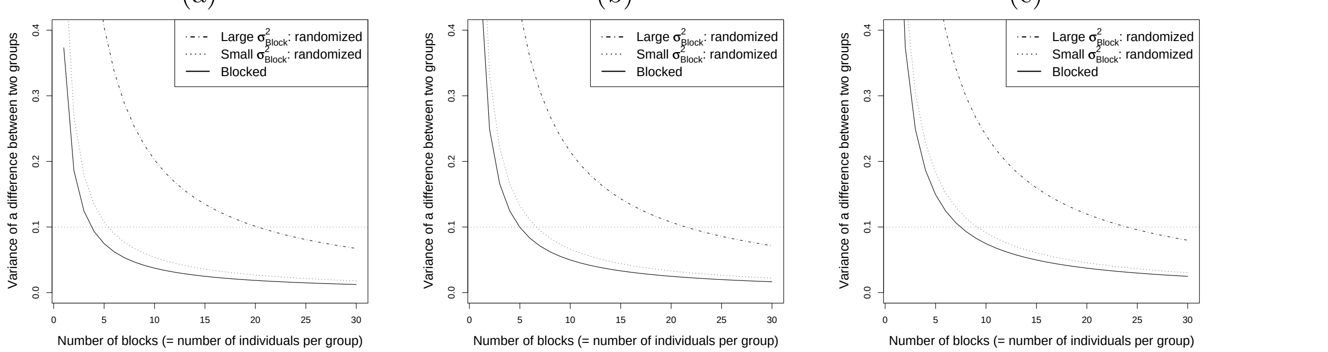

Figure 2: Blocking in a label-free workflow: extends Fig. 6(c) in the manuscript. (a) $\sigma^2_{Error} = 2\sigma^2_{Indiv}$; (b) $\sigma^2_{Error} = 3\sigma^2_{Indiv}$; (c) $\sigma^2_{Error} = 5\sigma^2_{Indiv}$.

Section 3.3.2: Blocking in a labeling workflow

The section considers the effect of blocking in a labeling design. The variances of a two-group comparisons of all but one design are functions of $\sigma^2_{Indiv} + \sigma^2_{Error}$ as in Sec. 3.3.1, thus an increase of σ^2_{Error} will have a similar effect. An exception is the reference design, for which the variance is

$$Var(\hat{D}_1 - \hat{D}_2) = \frac{2}{I}\left(\sigma^2_{Indiv} \ + 2\sigma^2_{Error}\right). \tag{4}$$

Fig. 3 illustrates the effect of an increasing σ^2_{Error} by extending Fig. 10(b) in the main manuscript. We keep σ^2_{Indiv} to the median experimental value as in the main manuscript, and increase σ^2_{Error}. One can see that (1) the increase in σ^2_{Error} results in an overall increase of the variance of the comparison, and (2) the increase of the variance in the reference design is slightly larger than for the other designs. While the variance of the reference design is similar to that of the balanced incomplete block design when $\sigma^2_{Error} = 2\sigma^2_{Indiv}$, it becomes closer to the variance of a disconnected pair of groups in a loop design when $\sigma^2_{Error} = 5\sigma^2_{Indiv}$.

Fig. 4 further illustrates the ef in order to fully compare various experimental settingsfect of increasing σ^2_{Error} in terms of the number of runs. The increase in σ^2_{Error} results in an overall increase in the variance of comparison given a number of runs, and the largest increase corresponds to the reference design.

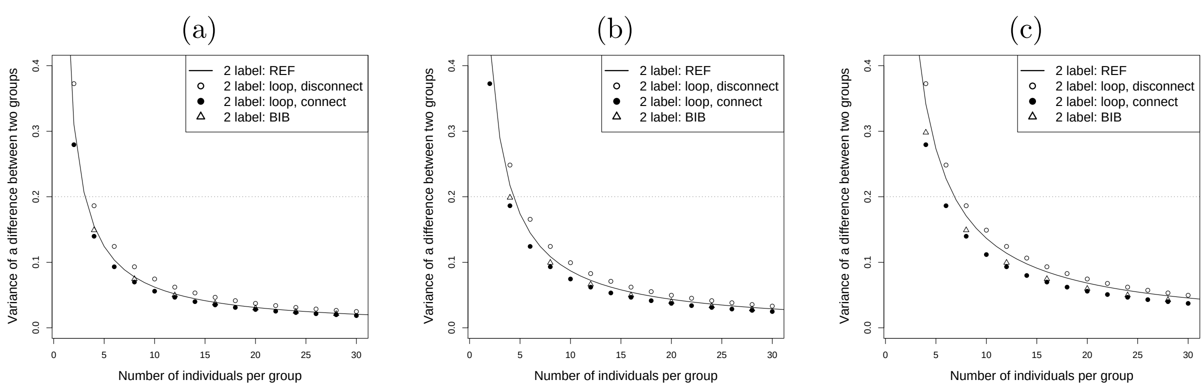

Figure 3: Blocking in a labeling workflow: extends Fig. 10(b) in the manuscript. (a) $\sigma^2_{Error} = 2\sigma^2_{Indiv}$; (b) $\sigma^2_{Error} = 3\sigma^2_{Indiv}$; (c) $\sigma^2_{Error} = 5\sigma^2_{Indiv}$.

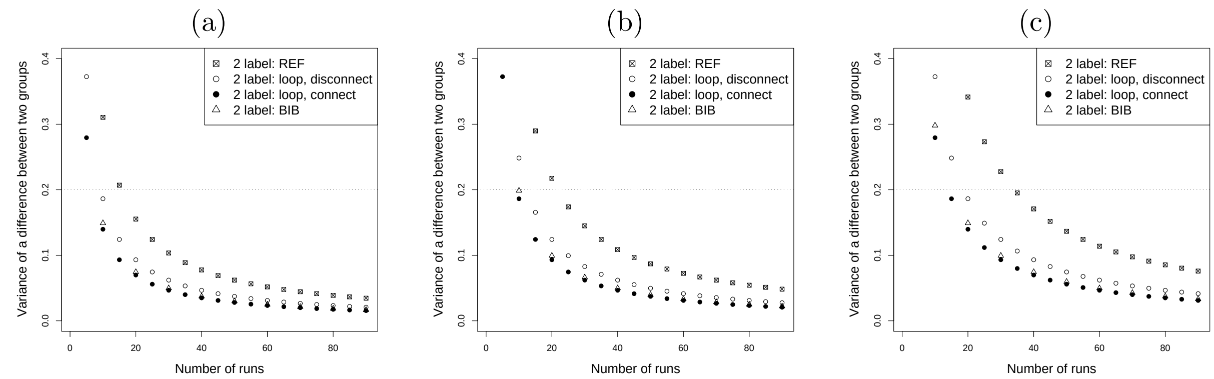

Figure 4: Blocking in a labeling workflow: extends Fig. 10(c) in the manuscript. (a) $\sigma^2_{Error} = 2\sigma^2_{Indiv}$; (b) $\sigma^2_{Error} = 3\sigma^2_{Indiv}$; (c) $\sigma^2_{Error} = 5\sigma^2_{Indiv}$.

Section 5.1: Sample size for a single feature

The variance components impact sample size calculations for a future experiment. When the variability increases, and the probability of Type I error and the power of the experiment are fixed, a larger number of replicates is necessary to detect a fold change. Fig. 5 extends Fig. 11(a) in the main manuscript and illustrates this effect by fixing σ^2_{Indiv} and σ^2_{Prep} to the median experimental values as in Fig. 11(a), and varying σ^2_{Error}.

One can see from the figure that (1) an increase in σ^2_{Error} results in an increase of the required sample size for all fold changes; (2) when the number of individuals is fixed and σ^2_{Error} increases, the relative effectiveness of technical replicates in detecting smaller fold changes also increases; and (3) when the total number of runs is fixed, allocating all the runs to the biological replicates remains advantageous in terms of detecting a smaller fold change. For example, in the panel (c) of the figure, an experiment without technical replicates which allocates 15 runs to 15 individuals in a group allows one to detect a smaller fold change than an experiment which allocates 15 runs to 5 individuals, 1 sample preparation and 3 technical replicates.

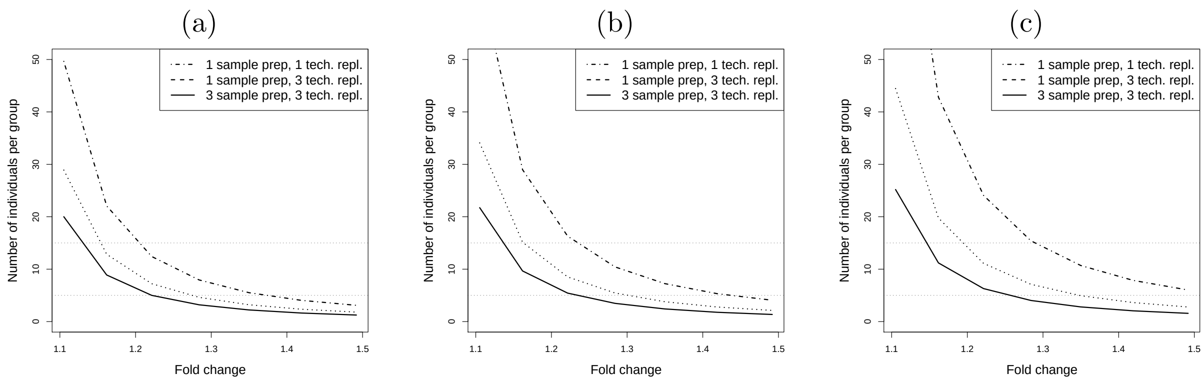

Figure 5: Blocking in a labeling workflow: extends Fig. 11(a) in the manuscript. (a) $\sigma^2_{Error} = 2\sigma^2_{Indiv}$; (b) $\sigma^2_{Error} = 3\sigma^2_{Indiv}$; (c) $\sigma^2_{Error} = 5\sigma^2_{Indiv}$.