

**In-depth method assessments of differentially expressed protein  
detection for shotgun proteomics data with missing values**

Jinxia Wang<sup>1,2\$\*</sup>, Liwei Li<sup>1\*</sup>, Tao Chen<sup>1</sup>, Jie Ma<sup>1</sup>, Yunping Zhu<sup>1</sup>, Jujuan Zhuang<sup>2</sup>,

Cheng Chang<sup>1</sup>

*<sup>1</sup>State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), National Engineering Research Center for Protein Drugs, Beijing Institute of Radiation Medicine, Beijing, 102206, P. R. China*

*<sup>2</sup>Department of Mathematics, Dalian Maritime University, Dalian, 116026, P.R. China.*

*<sup>\$</sup>Present address: Drug Research and Development Center, Shandong Drug and Food Vocational College, Weihai, 264210, P.R. China.*

*\*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to C.C. (email: 1987ccpacer@163.com) or J.Z. (email: jjzhuang@dlnu.edu.cn)*

## Supplementary Figures and Tables

### List of Figures

**Figure S1.** Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(b) with 10% missing value percentage.

**Figure S2.** Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(c) with 15% missing value percentage.

**Figure S3.** Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(d) with 20% missing value percentage.

**Figure S4.** Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(e) with 25% missing value percentage.

**Figure S5.** Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(f) with 30% missing value percentage.

**Figure S6.** Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(g) with 35% missing value percentage.

**Figure S7.** Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(h) with 40% missing value percentage.

percentage.

**Figure S8.** Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(i) with 45% missing value percentage.

### **List of Tables**

**Table S1.** Definitions of the indices used to evaluate the performances of DEP detection methods.

**Table S2.** Protein quantification results of experimental dataset D1 using MaxQuant.

**Table S3.** Protein quantification results of experimental dataset D2 using MaxQuant.

Figure S1. Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(b) with 10% missing value percentage.

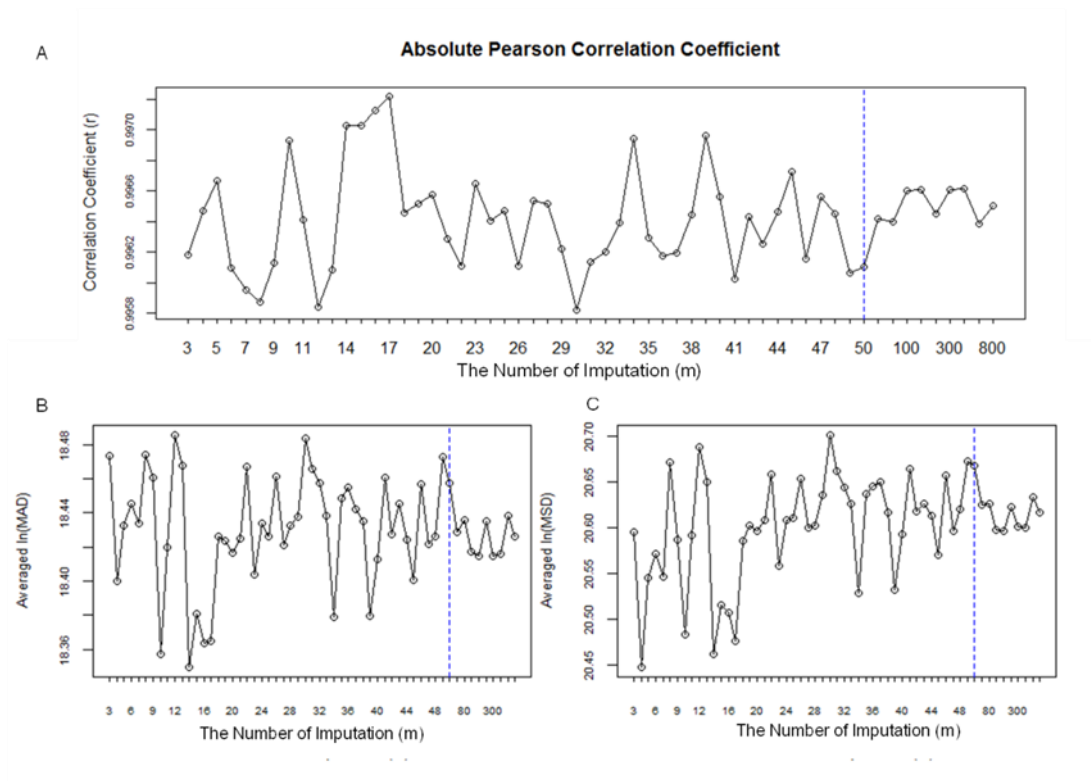


Figure S2. Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(c) with 15% missing value percentage.

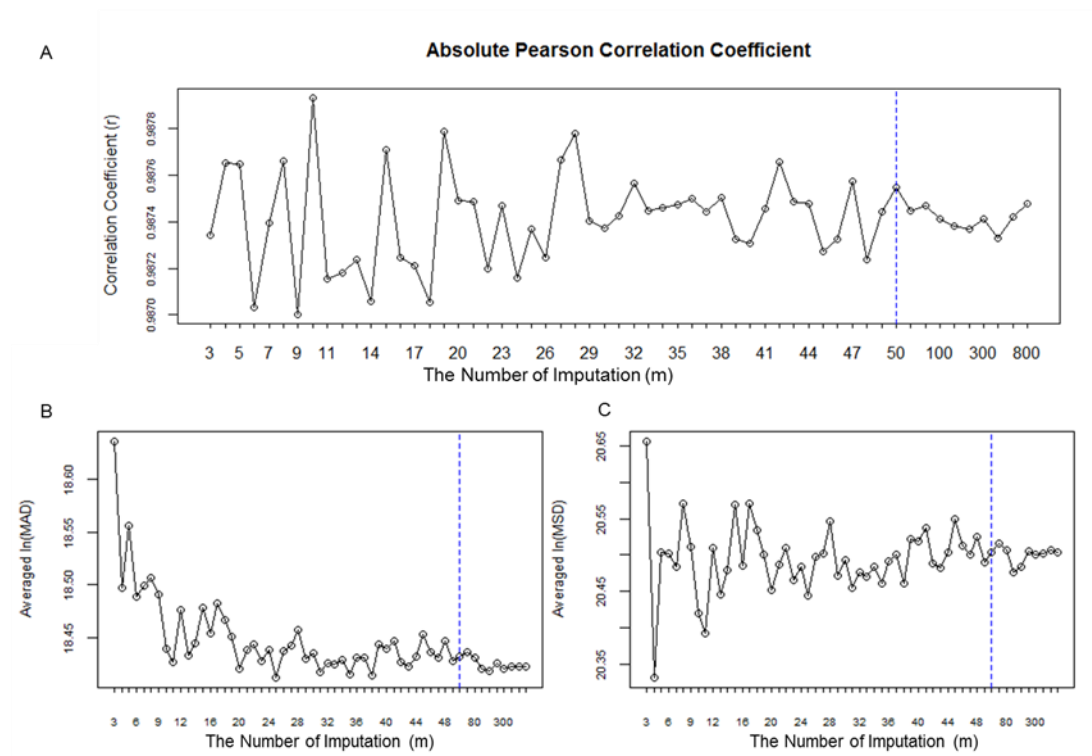


Figure S3. Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(d) with 20% missing value percentage.

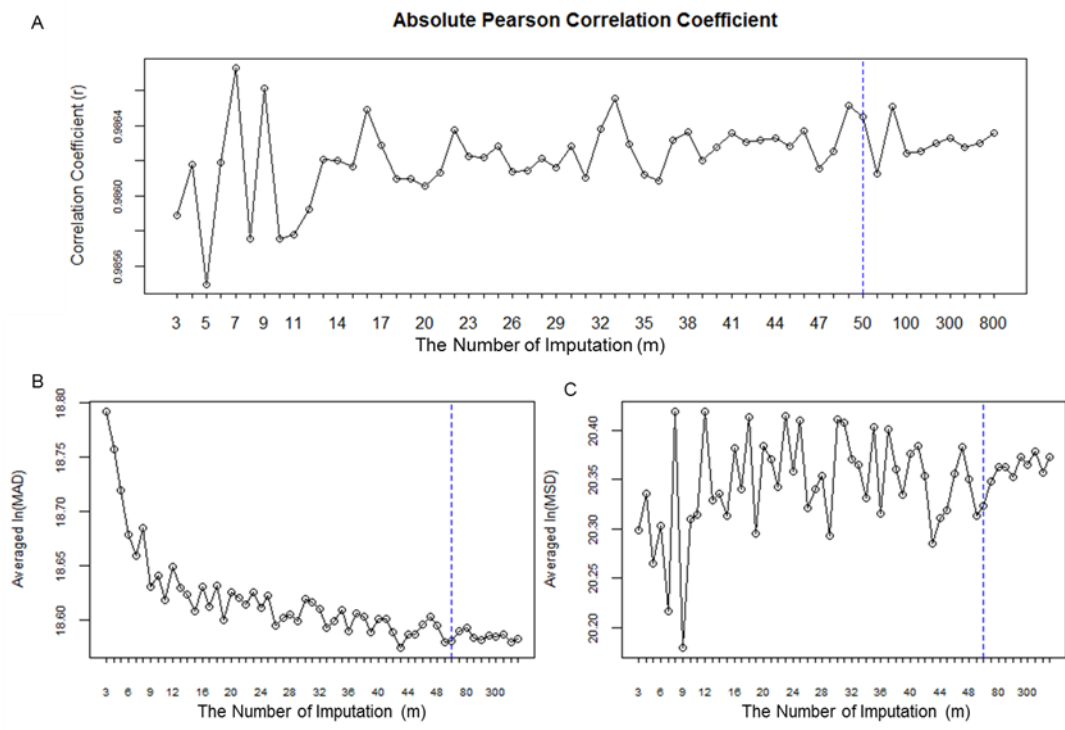


Figure S4. Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(e) with 25% missing value percentage.

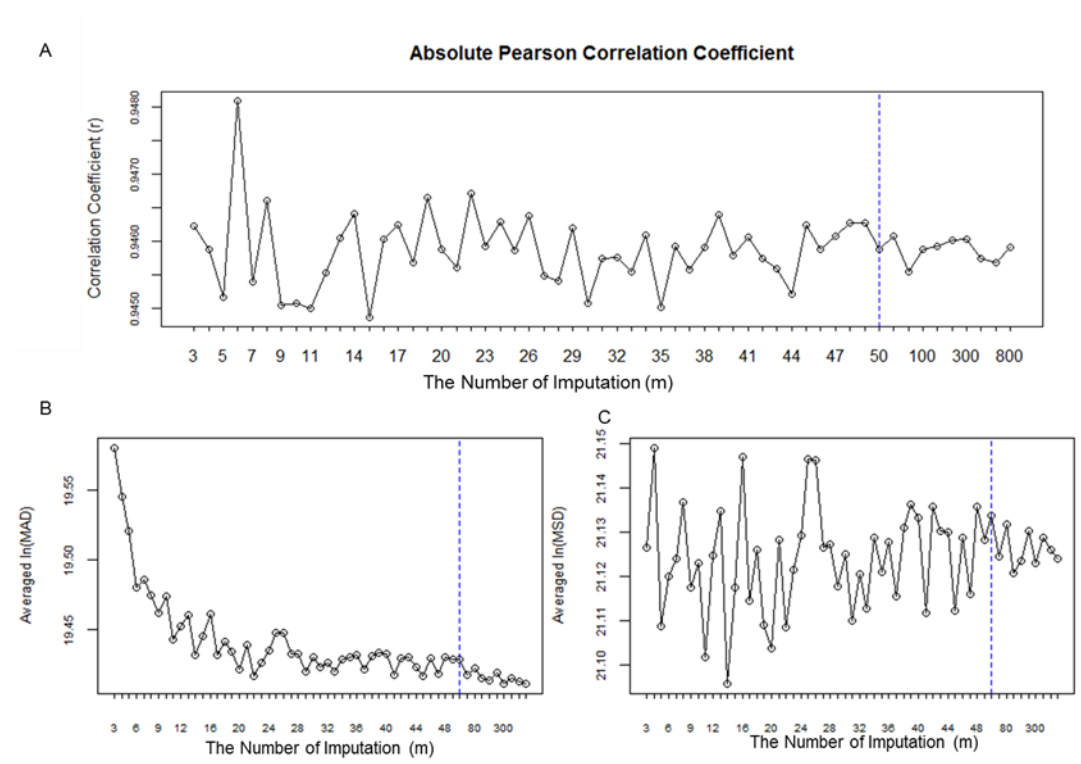


Figure S5. Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(f) with 30% missing value percentage.

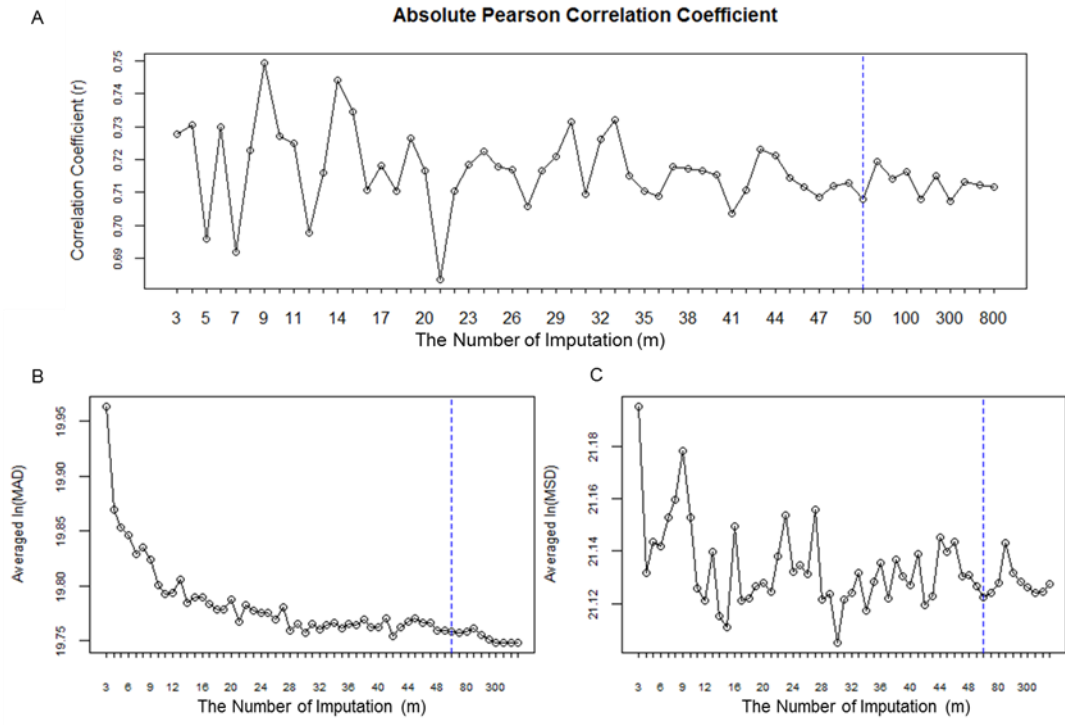




Figure S6. Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(g) with 35% missing value percentage.

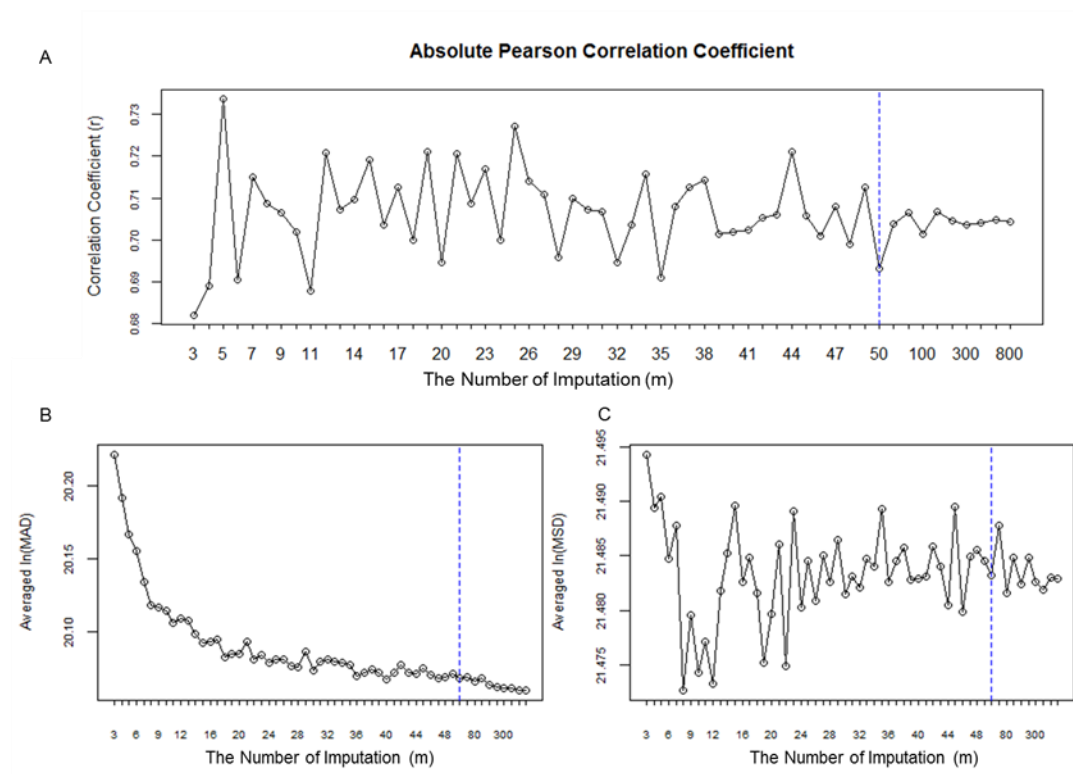


Figure S7. Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(h) with 40% missing value percentage.

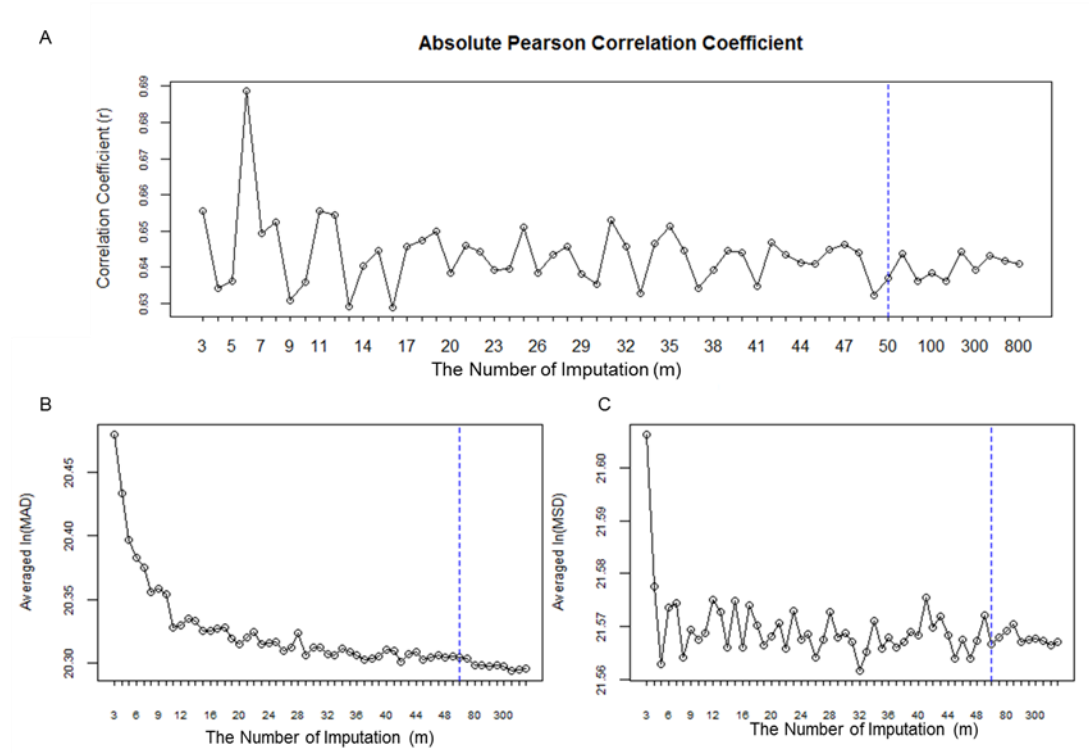


Figure S8. Line charts of absolute Pearson correlation coefficient as well as the averaged  $\ln(\text{MAD})$  and  $\ln(\text{MSD})$  in simulated dataset D0(i) with 45% missing value percentage.

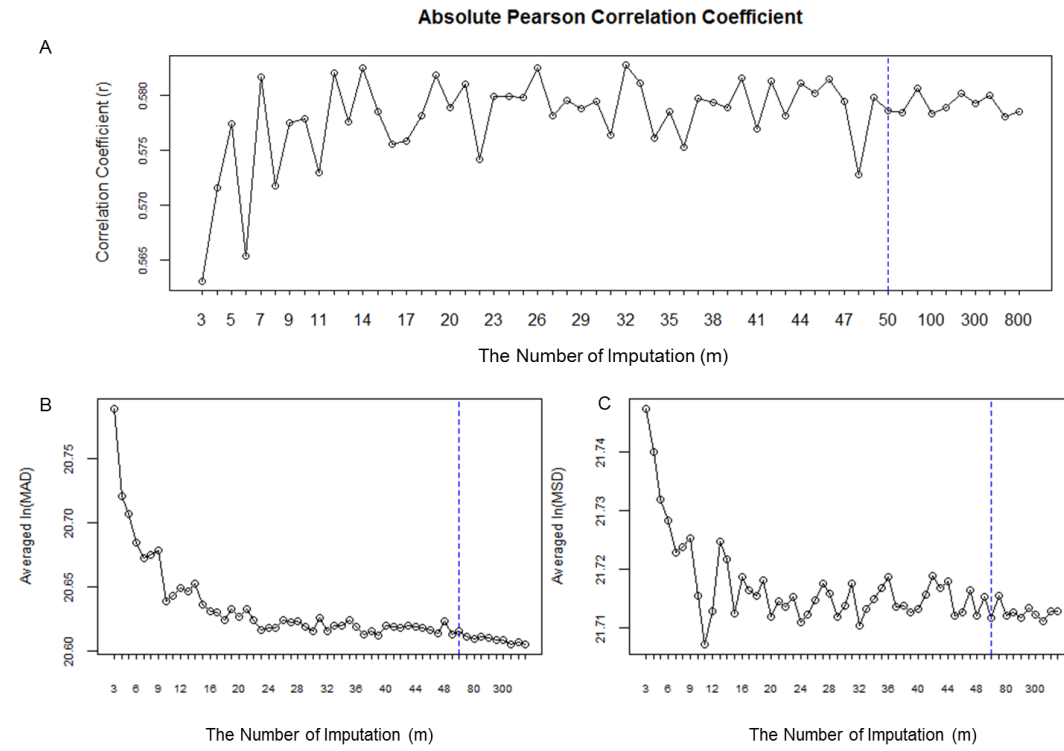


Table S1. Definitions of the indices used to evaluate the performances of DEP detection methods.

Measure	Equation
Sensitivity = Recall (R) = TPR	$\frac{TP}{TP + FN}$
Specificity = TNR = 1- FPR	$\frac{TN}{TN + FP}$
FDR	$\frac{FP}{FP + TP}$
Precision (P)	$\frac{TP}{TP + FP}$
f-score	$\frac{P \times R}{P + R}$
g-score (Geometric Mean Accuracy)	$\sqrt{TPR \times TNR}$