Medicine®

OPEN

A test for treatment effects in randomized controlled trials, harnessing the power of ultrahigh dimensional big data

Wen-Chung Lee, MD, PhD*, Jui-Hsiang Lin, MD, PhD

Abstract

Background: The randomized controlled trial (RCT) is the gold-standard research design in biomedicine. However, practical concerns often limit the sample size, *n*, the number of patients in a RCT. We aim to show that the power of a RCT can be increased by increasing *p*, the number of baseline covariates (sex, age, socio-demographic, genomic, and clinical profiles et al, of the patients) collected in the RCT (referred to as the ‘dimension’).

Methods: The conventional test for treatment effects is based on testing the ‘crude null’ that the outcomes of the subjects are of no difference between the two arms of a RCT. We propose a ‘high-dimensional test’ which is based on testing the ‘sharp null’ that the experimental intervention has no treatment effect whatsoever, for patients of any covariate profile.

Results: Using computer simulations, we show that the high-dimensional test can become very powerful in detecting treatment effects for very large *p*, but not so for small or moderate *p*. Using a real dataset, we demonstrate that the *P* value of the high-dimensional test decreases as the number of baseline covariates increases, though it is still not significant.

Conclusion: In this big-data era, pushing *p* of a RCT to the millions, billions, or even trillions may someday become feasible. And the high-dimensional test proposed in this study can become very powerful in detecting treatment effects.

Abbreviations: OC = operating characteristic, RCT = randomized controlled trial.

Keywords: big data, biostatistics, data mining, potential-outcome model, randomized controlled trial, sample size, sharp null

Strengths and limitations of this study

1. This paper presents a test for treatment effects in randomized controlled trials, which harnesses the power of ultrahigh dimensional big data.
2. The proposed high-dimensional test increases the power of a RCT by increasing p, the number of baseline covariates (sex, age, socio-demographic, genomic, and clinical profiles et al, of the patients), rather than the usual n, the number of patients.
3. The proposed high-dimensional test can become very powerful in detecting treatment effect for large p, but not so for small or moderate p.

1. Introduction

The randomized controlled trial (RCT) is the gold-standard research design in biomedicine and provides the most rigorous way of determining whether a cause-effect relation exists between treatment and outcome.[1,2] Randomization (random allocation of patients to intervention groups) and double blinding (neither the patients or investigators being aware of the treatment assignments

Editor: Jimmy T. Efird.

This paper is supported by grants from the Ministry of Science and Technology in Taiwan (MOST 105-2314-B-002-049-MY3, MOST 104-2314-B-002-118-MY3, MOST 108-3017-F-002-001, MOST 108-2314-B-002-127-MY3), and the Innovation and Policy Center for Population Health and Sustainable Environment (Population Health Research Center, PHRC) from Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan (NTU-108L9003).

The dataset supporting the conclusions of this article is included within the article and the supplementary files.

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan.

** Correspondence: Wen-Chung Lee, National Taiwan University, Rm. 536, No. 17, Xuzhou Rd., Taipei 100, Taiwan (e-mail: wenchung@ntu.edu.tw).*

How to cite this article: Lee WC, Lin JH. A test for treatment effects in randomized controlled trials, harnessing the power of ultrahigh dimensional big data. Medicine 2019;98:43(e17630).

Received: 1 October 2018 / Received in final form: 21 August 2019 / Accepted: 23 September 2019

http://dx.doi.org/10.1097/MD.0000000000017630

until the study is completed) are the hallmarks of RCTs. A carefully conducted RCT should be free from selection or confounding bias that otherwise plagues most observational studies.[3]

RCTs are, however, more costly and time-consuming than other studies. A realistic RCT is therefore often limited in sample size (n, the number of patients participating in the study) to no more than a few thousands patients. The power of a study is however an increasing function of n; an investigator content with a small n will likely get a non-significant result despite all the efforts he/she put into conducting the trial.[4] We are therefore posed with a dilemma—to recruit or not to recruit more patients.

We suggest new avenue for future RCTs. In this paper, we develop a "high-dimensional test" for treatment effects. We will show that the power of the test is an increasing function of p, the number of baseline covariates (sex, age, socio-demographic, genomic, and clinical profiles et al, of the patients) collected in a RCT (p is also referred to as the "dimension", and hence the name of the test). We will show that the high-dimensional test can become very powerful in detecting treatment effects if p can be made very large. We will also use a real dataset to demonstrate the methodology.

2. Methods

2.1. High-dimensional test

In a typical RCT comparing an experimental intervention and a suitable control for a continuous or binary end point, let A denote the treatment assignment indicator ($A=1$ for experimental intervention; $A=0$ for control), Y, the outcome, and $\mathbf{z}$, a vector of baseline covariates (with a dimension of p). We use the generic notation, $f(\cdot)$, to denote the (joint) probability density or mass function of a random variable (vector), where appropriate. The conventional test for treatment effects is based on testing the following 'crude null',

$$f(Y|A=1) = f(Y|A=0), \tag{1}$$

That is, the outcomes of the subjects are of no difference between the 2 arms of a RCT.

By contrast, the proposed high-dimensional test is based on testing the following "sharp null",

$$f(Y|A=1,\mathbf{z}) = f(Y|A=0,\mathbf{z}), \tag{2}$$

That is, the experimental intervention has no treatment affect whatsoever, for patients of any covariate profile. In practice, we can dichotomize Y into Y^*, such as 'favorable' ($Y^*=1$) and 'unfavorable' ($Y^*=0$) outcomes, based on some suitable criteria. (Y^* may already be a binary variable, such as 'survival' ($Y=1$) and 'death' ($Y=0$). This case is then simply $Y^*=Y$) Supplementary Note, http://links.lww.com/MD/D302 shows that alternatively, we can test the sharp null in a RCT, based on the following two equalities:

$$f(\mathbf{z}|A=1,Y^*=1) = f(\mathbf{z}|A=0,Y^*=1) \tag{3}$$

and

$$f(\mathbf{z}|A=1,Y^*=0) = f(\mathbf{z}|A=0,Y^*=0). \tag{4}$$

This alternative sharp-null formulation implies no difference in the baseline covariates between the two arms of a RCT, separately for those with favorable outcomes (3) and those with unfavorable outcomes (4).

Assume that a RCT recruits a total of n ($i=1,\ldots,n$) subjects. The data collected consists of the treatment assignment indicator, A_i, the outcome (and the dichotomized outcome), Y_i (and Y_i^*), and a total of p ($j=1,\ldots,p$) baseline covariates, Z_{ij}, for $i=1,\ldots,n$. To test the crude null (1), one can use the usual two-sample test,

$$T^2_{\text{crude}} = \left(\frac{\sum_{i:A_i=1} Y_i}{n_1} - \frac{\sum_{i:A_i=0} Y_i}{n_0}\right)^2 \Big/ \left(\frac{\hat{\sigma}^2_Y}{n_1} + \frac{\hat{\sigma}^2_Y}{n_0}\right), \tag{5}$$

where $n_1(n_0)$ is the number of subjects receiving the experimental (control) intervention ($n_1+n_0=n$), and $\hat{\sigma}^2_Y = \frac{1}{n-1} \times \sum_i \left(Y_i - \frac{1}{n} \times \sum_k Y_k\right)^2$ is an estimate of the variance of the outcome under the crude null. T^2_{crude} in (5) is distributed asymptotically as a chi-squared distribution with one degree of freedom under the crude null. The same can be done for the dichotomized outcome, Y_i^*.

To test the sharp null using (3) and (4), we can construct a test statistic for the jth baseline covariate,

$$\begin{aligned} T_j^2 = &\left(\frac{\sum_{i:A_i=1,Y_i^*=1} Z_{ij}}{n_{11}} - \frac{\sum_{i:A_i=0,Y_i^*=1} Z_{ij}}{n_{01}}\right)^2 \Big/ \left(\frac{\hat{\sigma}^2_{j,1}}{n_{11}} + \frac{\hat{\sigma}^2_{j,1}}{n_{01}}\right) \\ &+ \left(\frac{\sum_{i:A_i=1,Y_i^*=0} Z_{ij}}{n_{10}} - \frac{\sum_{i:A_i=0,Y_i^*=0} Z_{ij}}{n_{00}}\right)^2 \Big/ \left(\frac{\hat{\sigma}^2_{j,0}}{n_{10}} + \frac{\hat{\sigma}^2_{j,0}}{n_{00}}\right), \end{aligned} \tag{6}$$

where $n_{11}(n_{01})$ and $n_{10}(n_{00})$ are the numbers of subjects receiving the experimental (control) intervention and ultimately leading to, respectively, favorable and unfavorable outcomes ($n_{11}+n_{01}+n_{10}+n_{00}=n$), and $\hat{\sigma}^2_{j,1} = \frac{1}{n_{11}+n_{01}-1} \times \sum_{i:Y_i^*=1} \left(Z_{ij} - \frac{1}{n_{11}+n_{01}} \times \sum_{k:Y_k^*=1} Z_{kj}\right)^2$ and $\hat{\sigma}^2_{j,0} = \frac{1}{n_{10}+n_{00}-1} \times \sum_{i:Y_i^*=0} \left(Z_{ij} - \frac{1}{n_{10}+n_{00}} \times \sum_{k:Y_k^*=0} Z_{kj}\right)^2$ are the estimates of the variances of the jth baseline covariate under the sharp null among subjects with, respectively, favorable and unfavorable outcomes. The first term to the right of the equality sign in (6) is a test statistic based on (3), and the second term, that based on (4). These 2 terms involve different sets of subjects and are independent of one another. Under the sharp null, T_j^2 in (6) is therefore distributed asymptotically as a chi-squared distribution with 2 degrees of freedom.

Next, we sum up the statistics of all p baseline covariates as our high-dimensional test,

$$T^2_{\text{sharp}} = \sum_{j=1}^{p} T_j^2. \tag{7}$$

The ordinary chi-square approximation may not apply for T^2_{sharp} in (7) because the baseline covariates themselves may not be independent of one another. We, therefore, propose performing Monte-Carlo permutations for the sampling distribution of T^2_{sharp} under the sharp null. To be precise, we fix $\mathbf{z}$ and shuffle (A, Y^*) among the study subjects (or vice versa). The permutation-based high-dimensional test is a distribution-free test, suitable for use with normal or non-normal data and in large or small RCTs.

2.2. Simulation study

We considered a small RCT with $n=50$ and a large one with $n=250$. Each patient is randomized either to the treatment or the control arm, with equal probability. The outcomes of the trials (survival or death) are recorded for each patient. The trials also collected p baseline covariates for each patient.

We assume a potential-outcome model[3,5] for a particular disease: the experimental treatment is beneficial to 15% of patients (they will live upon being given the treatment and will die otherwise), is harmful to 5% of patients (they will instead die upon being given the treatment but will live otherwise), and is of absolutely no effect on the rest (30% and 50% of patients are destined to live or die, respectively, regardless of the treatment given). We also consider a stochastic version of the model, in which those who will live or die *as per* the above deterministic model will succumb to the same fates, not absolutely but with a probability of 0.9. To check the validity of the high-dimensional test, we construct a sharp null of a deterministic potential-outcome model where no one is responsive to the treatment (assuming 40% patients are destined to live, and the other 60% will die, regardless of the treatment).

We assume that the baseline covariates are normally distributed with a constant variance of one, but with slightly different means for subjects of different potential-outcome types. To be precise, the type-specific means are randomly sampled from a $N(0,\Delta^2)$ normal distribution. In the simulation, we consider 3 scenarios for the association between the measured baseline covariates and the assumed potential-outcome types: (i) weak-to-moderate association ($\Delta^2=0.03$), (ii) weak association ($\Delta^2=0.01$), and (iii) ultra-weak association ($\Delta^2=0.005$). The baseline covariates are assumed to be independent of one another conditional on the potential-outcome types. We also considered the cases of weakly and strongly correlated covariates, where the correlation coefficients between the ith and the jth baseline covariates are assumed to be $0.5^{|i-j|}$ and $0.9^{|i-j|}$,respectively.

We simulate a hypothetical omniscient test to serve as an upper bound for what a real-world high-dimensional test can achieve. To be precise, an omniscient trial analyst having the knowledge regarding the potential-outcome types of all patients and puts this piece of information into the analysis; he/she creates four indicator variables, each indicating whether a subject belongs to a specific potential-outcome type, and then calculates a high-dimensional test treating these indicator variables as four "baseline covariates".

The "operating characteristic" (OC) of a test is its statistical power averaged over a uniformly distributed α-level between 0 and 1. The OC is a value between 0.5 (no power at all) and 1 (highest power possible). It can be converted to a power at a specified α-level, if the test statistic is normally distributed: $1-\Phi(Z_{1-\alpha/2}-\delta)+\Phi(Z_{\alpha/2}-\delta)$, where $\delta=\sqrt{2}\times Z_{\mathrm{OC}}$, and $\Phi(\cdot)$ is the cumulative distribution function, and Z_x, the x'th quantile of the standard normal distribution. In the simulation study, the OC is estimated as the proportion of the simulations that result in a test statistic larger than the same statistic under a random permutation of the data (as described before). If a test statistic happens to be equal to its permuted counterpart, a 0.5 count is tallied. A total of 1000 simulations were performed for each sharp-alternative scenario. To facilitate comparison, we estimated the OCs of the above omniscient test and the traditional test (testing the crude null) using the same simulation-permutation scheme as we used for the high-dimensional test.

For each sharp-null scenario, we performed a total of 10,000 simulations to estimate the OC and the type I error rate at $\alpha=0.05$ (with 99 permutations to derive the null sampling distribution in each round of the simulation).

2.3. Real data analysis

We re-analyzed Gene Expression Omnibus dataset (GSE118657) to illustrate the methodology.[6] The dataset is a Phase II randomized controlled trial assessing the effect of lactoferrin on critically ill patients undergoing mechanical ventilation (a total of 61 patients, 32 patients in the lactoferrin group, and the remaining, the placebo group). Gene expressions with a total of 49,495 genes were measured at the first day of admission for each patient. The proposed high-dimensional test was used to test the effect of lactoferrin treatment using all gene expressions as the baseline covariates ($p=49{,}495$). We also examined the effects of using reduced numbers of genes ($p=1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10{,}000, 20{,}000$, respectively) randomly sampled from the total 49,495 genes (100 random samples were taken and the results were averaged for each scenario). A total of 9999 permutations were performed to derive the null sampling distribution for the high-dimensional test.

2.4. Ethical review

This paper is a methodological study (computer simulation study) and does not involve the enrollment of patients. The real data used in this paper is from public domain. Ethical approval is not necessary.

3. Results

3.1. Simulation study

Figure 1 presents the results when the outcomes follow the assumed deterministic potential-outcome model. For a small RCT ($n=50$), the traditional test (testing the crude null) has a very low OC of 0.57, whereas the omniscient test can have a very high OC of 0.93. When the sample size increases to $n=250$, the performance of the traditional test improves, though not by very much (OC=0.78), whereas the omniscient test now functions impeccably (OC=1.00).

In a real-world RCT, the potential-outcome types of the patients are, of course, unknown. However, we found that the performance of the hypothetical omniscient test can be replicated using a real-world high-dimensional test (Fig. 1). With a large enough p (more than 10 weak-to-moderate covariates, more than 100 weak covariates, or more than 1000 ultra-weak covariates), the high-dimensional test outperforms the traditional test. For a large RCT (such as when $n=250$) and with a fairly large p (such as when $p>10^4$), the high-dimensional test can also become impeccable (OC→1).

The high-dimensional test is, as it should be, bounded above by the omniscient test in terms of its OC, no matter how strong the association is between the covariates used and the potential-outcome types, and no matter how many there are (Supplementary Fig. 1, http://links.lww.com/MD/D302).

Under the sharp null, the high-dimensional test has an OC close to 0.5 (Table 1) and a type I error rate close to the nominal α level of 0.05 for all scenarios studied (Table 2).

Figure 2 presents the results when the outcomes follow the stochastic potential-outcome model. Again, we see that the

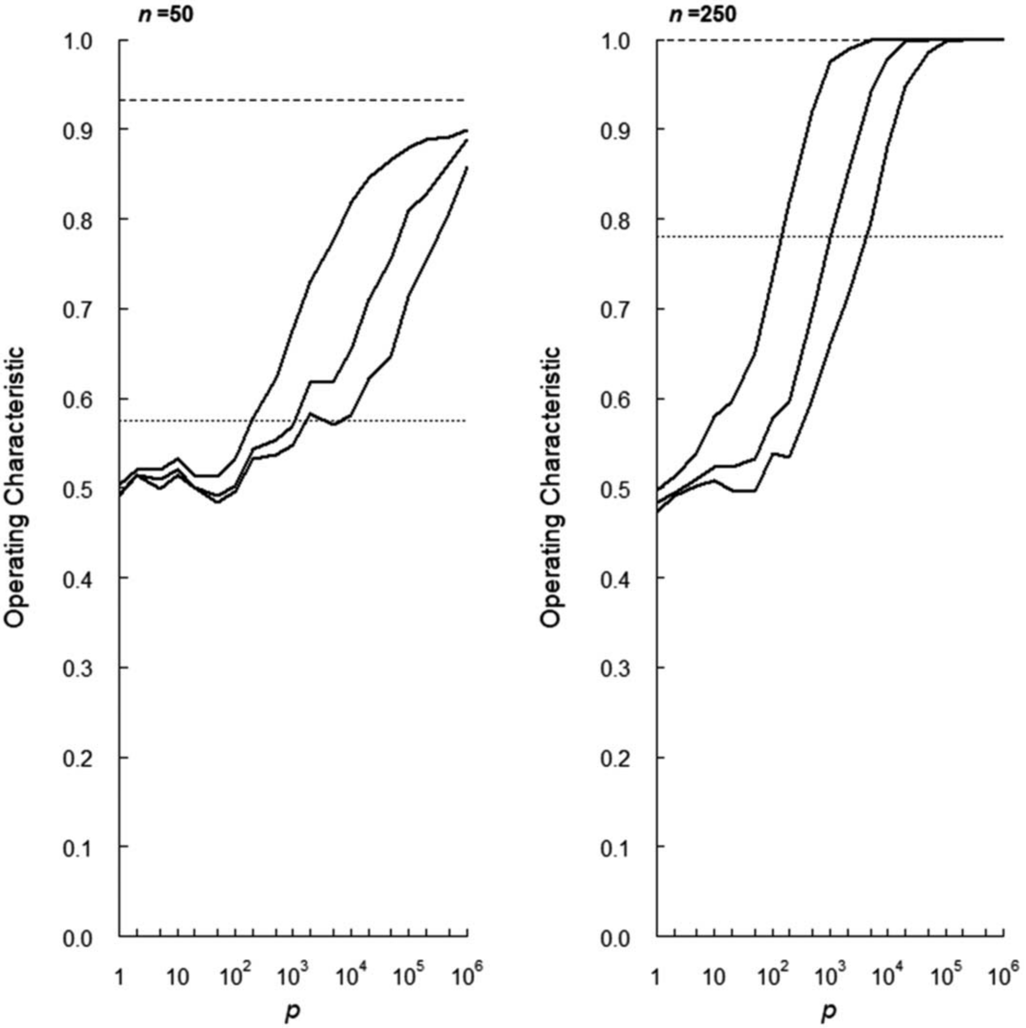

Figure 1. Operating characteristics of the traditional test (dotted horizontal lines), the high-dimensional test (left solid curves: weak-to-moderate covariates; middle solid curves: weak covariates; right solid curves: ultra-weak covariates), and the omniscient test (dash horizontal lines), under a deterministic potential-outcome model.

traditional test performs very poorly (OC=0.55 when $n=50$; OC=0.67 when $n=250$). With the stochasticity introduced, a perfect knowledge of the potential-outcome types no longer foretells a subject's fate exactly (only with an accuracy rate of 0.9 for the assumed model). Yet, the omniscient test still performs remarkably better than the traditional test in a small trial (OC=0.84 when $n=50$), and can even become impeccable in a large RCT (OC=1.00 when $n=250$).

Again, the (real-world) high-dimensional test outperforms the traditional test with a large enough p (Fig. 2). It can also become

Table 1

Operating characteristics of the high-dimensional test under the sharp null.

Number of subjects (n) and strength of the covariates	Number of baseline covariates (p)												
	1	2	5	10	20	50	100	200	500	1000	2000	5000	10000
$n=50$													
weak to moderate	0.5038	0.5008	0.4993	0.4947	0.4990	0.4977	0.4974	0.4992	0.4943	0.4981	0.5001	0.5024	0.4977
weak	0.4972	0.5001	0.5064	0.5087	0.5048	0.5045	0.5016	0.5048	0.5047	0.5057	0.4984	0.4911	0.4918
ultra weak	0.5014	0.5069	0.4974	0.4961	0.5004	0.5068	0.4983	0.5058	0.5050	0.5062	0.4985	0.5048	0.5019
$n=250$													
weak to moderate	0.5005	0.4982	0.4952	0.5017	0.5069	0.5158	0.5072	0.5020	0.5017	0.5055	0.5021	0.4877	0.4949
weak	0.5033	0.5000	0.4987	0.5021	0.5043	0.4991	0.4967	0.5009	0.4943	0.4990	0.5095	0.4978	0.5000
ultra weak	0.4998	0.5067	0.5078	0.5017	0.4980	0.4973	0.4965	0.4962	0.4974	0.4957	0.4906	0.4989	0.4999

Table 2

Type I error rates at $\alpha = 0.05$ of the high-dimensional test under the sharp null.

Number of subjects (n) and strength of the covariates	Number of baseline covariates (p)												
	1	2	5	10	20	50	100	200	500	1000	2000	5000	10000
$n=50$													
weak to moderate	0.0451	0.0470	0.0496	0.0498	0.0519	0.0503	0.0518	0.0493	0.0485	0.0495	0.0540	0.0488	0.0488
weak	0.0532	0.0502	0.0488	0.0465	0.0459	0.0475	0.0514	0.0493	0.0543	0.0534	0.0529	0.0485	0.0510
ultra weak	0.0498	0.0485	0.0516	0.0507	0.0490	0.0508	0.0513	0.0524	0.0507	0.0484	0.0506	0.0475	0.0490
$n=250$													
weak to moderate	0.0514	0.0509	0.0498	0.0494	0.0503	0.0518	0.0522	0.0558	0.0536	0.0465	0.0527	0.0461	0.0469
weak	0.0506	0.0464	0.0463	0.0513	0.0495	0.0528	0.0511	0.0481	0.0511	0.0519	0.0506	0.0495	0.0503
ultra weak	0.0477	0.0499	0.0484	0.0489	0.0517	0.0484	0.0456	0.0475	0.0527	0.0484	0.0469	0.0485	0.0502

impeccable in a large RCT ($n=250$) with $p>10^6$, or with a smaller p if the covariates used are more strongly associated with the potential-outcome types (Supplementary Fig. 2, http://links.lww.com/MD/D302).

Table 3 compares the OCs of the high-dimensional test for independent, weakly correlated, and strongly correlated, baseline covariates. With the same number of baseline covariates, the operating characteristic is lower if the baseline covariates are correlated with one

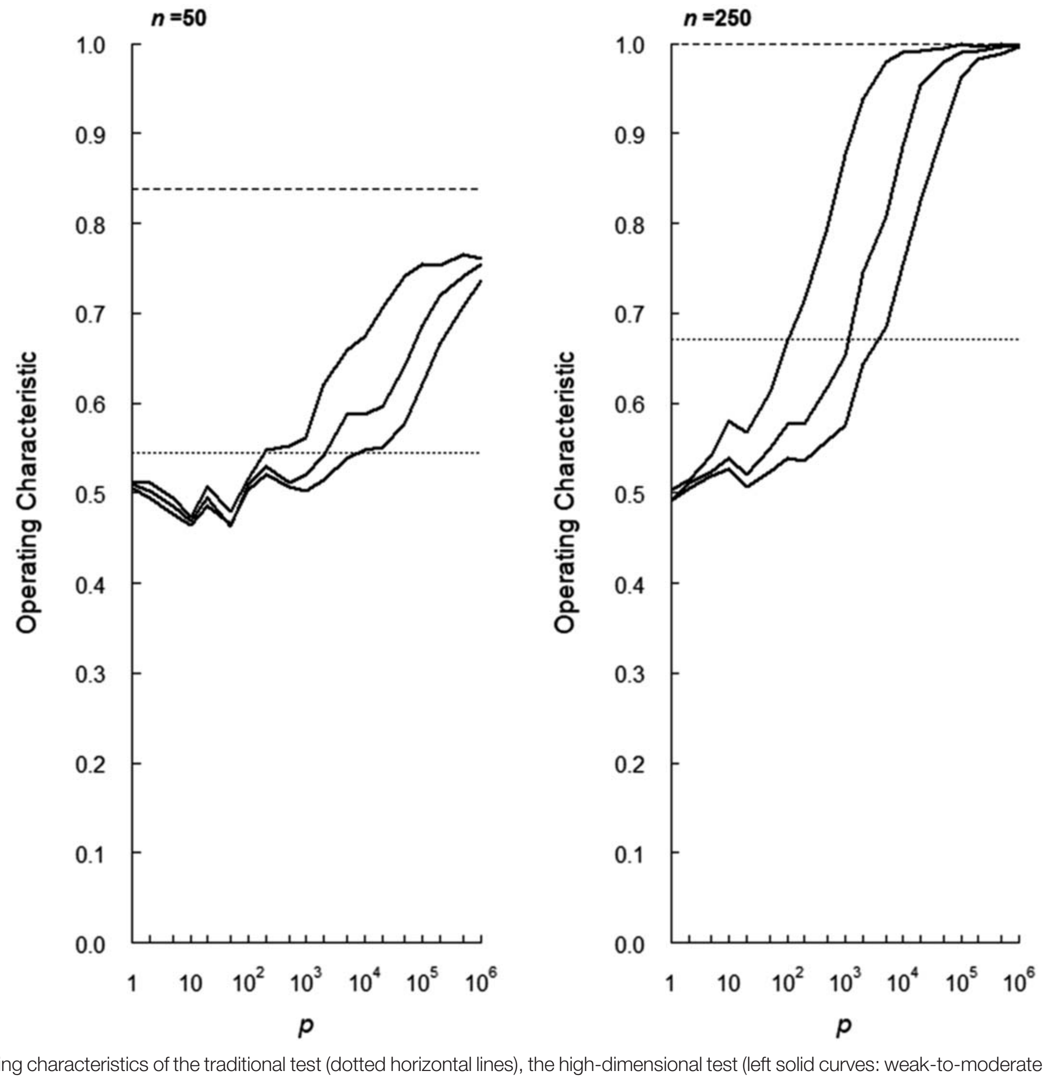

Figure 2. Operating characteristics of the traditional test (dotted horizontal lines), the high-dimensional test (left solid curves: weak-to-moderate covariates; middle solid curves: weak covariates; right solid curves: ultra-weak covariates), and the omniscient test (dash horizontal lines), under a stochastic potential-outcome model.

Table 3

Operating characteristics of the high-dimensional test for independent, weakly correlated, and strongly correlated, baseline covariates ($n=250$; strength of the covariates: weak to moderate).

	Number of baseline covariates (p)					
Correlation between covariates	**20**	**50**	**100**	**200**	**500**	**1000**
Deterministic potential-outcome model						
independent	0.595	0.685	0.756	0.826	0.929	0.976
weakly correlated	0.604	0.631	0.686	0.772	0.851	0.920
strongly correlated	0.563	0.577	0.613	0.660	0.710	0.783
Stochastic potential-outcome model						
independent	0.556	0.625	0.654	0.728	0.826	0.883
weakly correlated	0.559	0.574	0.606	0.650	0.751	0.826
strongly correlated	0.545	0.543	0.567	0.576	0.636	0.678

another. To make up for the power loss in using correlated covariates, one can include more covariates in the high-dimensional test. For all scenarios studied, OC increases as p increases.

We also performed additional simulations for more complexly distributed baseline covariates (non-normal covariates, a mixed panel of binary and continuous variables, and a mixed panel of signals and noises, see Supplementary Table, http://links.lww.com/MD/D302), and for a patient population with a different potential-outcome-type distribution from that assumed in this study (including ‘monotonicity’ scenarios where the experimental treatment can do only good and no harm[7]). The basic conclusions are the same though some scenarios may call for a larger p to achieve the same OC as in this paper.

However, the high-dimensional test has no power whatsoever to test the sharp null if none of the baseline covariate collected is a signal, or if the signal-to-noise ratio tends to zero as p tends to infinity. The high-dimensional test is also ineffective if the treatment effect is homogeneous across covariate profiles [e.g., all patients are of the same stochastic potential-outcome type: they all have the same survival probabilities of, say, 0.7(0.4), if given (not given) the treatment].

3.2. Real data analysis

Figure 3 presents the P values for the lactoferrin treatment. The traditional test (testing the crude null) has a P value of .36. As the number of baseline covariates (genes) increases, the P values of the high-dimensional test decreases. With 20 genes used, the high-dimensional test has a P value of .32, which is smaller than that of the traditional test. With all 49,495 genes used, the high-dimensional test has a P value of .23, though it is still not significant. To achieve significance (if the sharp null is indeed false for this example), one could include more baseline covariates into the high-dimensional test for the total 61 patients in the trial (as the power of the test is an increasing function of the number of covariates), and ideally covariates of diverse types other than the gene expression data currently used (as the power of the test is compromised for highly correlated covariates such as gene expressions).

4. Discussion

The proposed high-dimensional test is based on testing the sharp null. The sharp-null formulation in (2) is self-explanatory: the experimental intervention has no treatment affect whatsoever, for patients of any covariate profile. However, the sharp-null formulation in (3) and (4) seems rather peculiar. A simple two-step conditionality argument (Supplementary Note, http://links.lww.com/MD/D302) may help clarify what this alternative formulation means: (the first step) it is true that there shall be no association *unconditionally* between treatment assignment and each and every baseline covariate in a dutifully conducted RCT

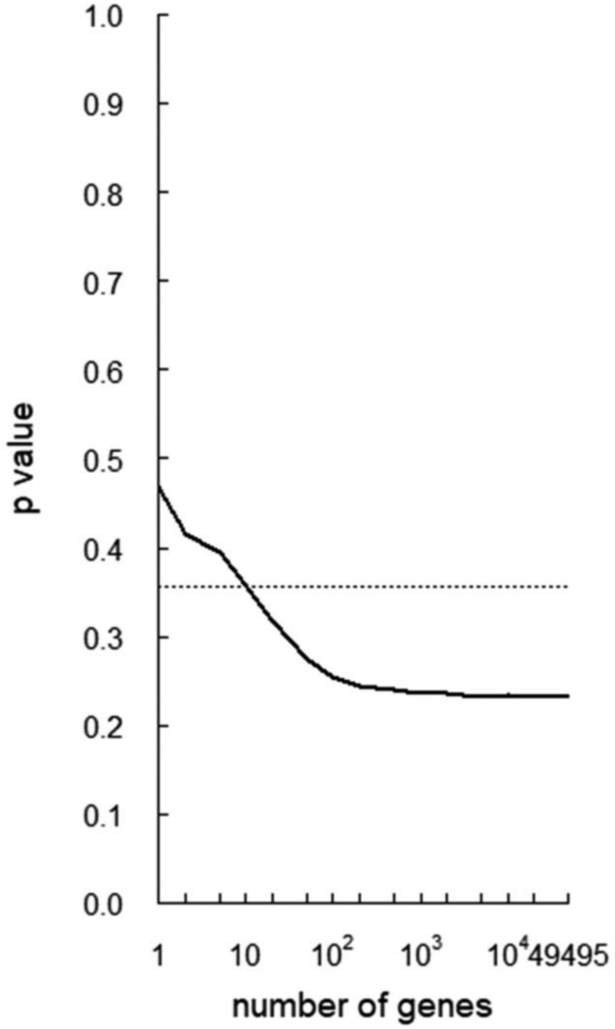

Figure 3. P values in GEO118657 dataset analysis (high-dimensional test: solid line; traditional test: dotted horizontal line).

($A \perp \mathbf{z}$, where the $\perp$ sign indicates 'independence' or 'no association'), and (the second step) if the sharp null in (2) is also true ($A \perp Y|\mathbf{z}$), then (the result) there shall furthermore be no association between treatment assignment and each and every baseline covariate, *conditional on the outcome* (the alternative sharp-null formulation, $A \perp \mathbf{z}|Y$).

Conventional wisdom holds that testing many variables simultaneously incurs a penalty[8] and many researchers turn to dimension reduction methods to mitigate the problem.[9–12] The "p" -based methods developed by previous researchers approached this multiple-testing problem differently, whereby the dimensionality is no longer a curse but in fact a blessing. For examples, Hall et al[13] and Ahn et al[14] studied the geometric properties of high-dimension and low-sample-size data and showed that the group memberships of study subjects can be resolved almost perfectly using their pair-wise distances (in high dimension), and Lo and Lee[15] constructed a p-based test to detect weak associations (when p is very large) and Lee[16] further developed a p-based adjustment method to correct for unmeasured confounding biases (again, when p is very large). In this paper, we extend the applicability of the "p"-based approach to RCT settings and show that the high-dimensional test can become very powerful in detecting treatment effects for very large p, the number of baseline covariates.

The current practice of RCTs follows the "n" -based paradigm; the power of a test is gauged by n, the number of study subjects.[4] But this has a limit as the n is bounded above by the world population. By contrast, in this big-data era[17–19] pushing the p of a RCT to the billions, trillions or even more may quickly become possible. The high-dimensional test we proposed in this paper thus provides a means to break the "n"-barrier and let ultrahigh dimensional big data generate new knowledge. But one needs to keep in mind that RCTs often have stringent inclusion and exclusion criteria. Even if infinite number of baseline covariates was collected in a RCT, the results of the high-dimensional test only apply to the (selected) patient population of that particular RCT and are not directly generalizable to patients seen in real-world.

For small or moderate p, say, hundreds, thousands or millions, the high-dimensional test by itself may be underpowered and should best be used in conjunction with the traditional test. A possible solution is to combine the "p" -based sharp-null test in (7) and the "n" -based crude-null test in (5): $w_{\text{sharp}} \times T^2_{\text{sharp}} + w_{\text{crude}} \times T^2_{\text{crude}}$, where w_{sharp} and w_{crude} are the weights attached, respectively, to the 2 tests. Further work is needed to study how to set the weights and to examine the statistical properties of this combined test. From our simulation study, the power of the high-dimensional test depends on many factors: the number of baseline covariates, the number of study subjects, the strength of the association between the baseline covariates and the potential-outcome types, the nature of the potential outcomes (deterministic or stochastic), the degree of the correlation between the baseline covariates, the distribution of the baseline covariates, the distribution of the potential-outcome types, etc. Further work is also needed to develop power formula for the proposed high-dimensional test.

5. Conclusions

In this big-data era, pushing p of a RCT to the millions, billions, or even trillions may someday become feasible. And the high-dimensional test proposed in this study can become very powerful in detecting treatment effects.

Author contributions

Conceptualization: Wen-Chung Lee.
Data curation: Wen-Chung Lee, Jui-Hsiang Lin.
Formal analysis: Wen-Chung Lee, Jui-Hsiang Lin.
Funding acquisition: Wen-Chung Lee.
Investigation: Wen-Chung Lee.
Methodology: Wen-Chung Lee.
Project administration: Wen-Chung Lee.
Resources: Wen-Chung Lee.
Software: Wen-Chung Lee, Jui-Hsiang Lin.
Supervision: Wen-Chung Lee.
Validation: Wen-Chung Lee.
Visualization: Wen-Chung Lee.
Writing – original draft: Wen-Chung Lee.
Writing – review & editing: Wen-Chung Lee.

References

[1] Sibbald B, Roland M. Why are randomised controlled trials important? Br Med J 1998;**316**:201.
[2] Kao LS, Tyson JE, Blakely ML, et al. Clinical research methodology I: introduction to randomized trials. J Am Coll Surg 2008;206:361–9.
[3] Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott; 2008.
[4] Lachin JM. Introduction to sample size determination and power analysis for clinical trials. Cont Clin Trials 1981;2:93–113.
[5] Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. J Am Stat Assoc 2005;100:322–31.
[6] Muscedere J, Maslove DM, Boyd JG, et al. Prevention of nosocomial infections in critically ill patients with lactoferrin: a randomized, double-blind, placebo-controlled study. Crit Care Med 2018;46:1450–6.
[7] Chiba Y. Bounds on causal effects in randomized trials with noncompliance under monotonicity assumptions about covariates. Stat Med 2009;28:3249–59.
[8] Noble WS. How does multiple testing correction work? Nat Biotechnol 2009;27:1135–7.
[9] Tibshirani R. Regression shrinkage and selection via the Lasso. J Royal Stat Soc (series B) 1996;58:267–88.
[10] Mehmood T, Liland KH, Snipen L, et al. A review of variable selection methods in partial least squares regression. Chemometr Intell Laborat Syst 2012;118:62–9.
[11] Fukumizu K, Bach FR, Jordan MI. Kernel dimension reduction in regression. Ann Stat 2009;37:1871–905.
[12] Hung H, Wang CC. Matrix variate logistic regression model with application to EEG data. Biostatistics 2013;14:189–202.
[13] Hall P, Marron JS, Neeman A. Geometric representation of high dimension, low sample size data. J Royal Stat Soc (series B) 2005;67: 427–44.
[14] Ahn J, Marron JS, Muller KM, et al. The high-dimension, low-sample-size geometric representation hold under mild conditions. Biometrika 2007;94:760–6.
[15] Lo MT, Lee WC. Detecting a weak association by testing its multiple perturbations: a data mining approach. Sci Rep 2014;4:5081.
[16] Lee WC. Detecting and correcting the bias of unmeasured factors using perturbation analysis: a data-mining approach. BMC Med Res Methodol 2014;14:18.
[17] Murdoch TB, Detsky AS. The inevitable application of big data to health care. J Am Med Assoc 2013;309:1351–2.
[18] Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Affairs 2014;33:1163–70.
[19] Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med Genomics 2015;8:33.