

# On the Application of Bootstrapping and Monte Carlo Simulations to Clinical Studies: Psychometric Intelligence Research and Juvenile Delinquency

Kohske Ogata 

Faculty of Human and Social Science, Osaka Ohtani University, Tondabayashi, Japan  
Email: koh-ske@sakai.zaq.ne.jp

**How to cite this paper:** Ogata, K. (2021). On the Application of Bootstrapping and Monte Carlo Simulations to Clinical Studies: Psychometric Intelligence Research and Juvenile Delinquency. *Psychology*, 12, 1171-1183.  
<https://doi.org/10.4236/psych.2021.128072>

**Received:** July 9, 2021  
**Accepted:** July 30, 2021  
**Published:** August 3, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The common problems in the methodology of clinical psychology research are sampling issues, both in the case of biased clinical groups and inappropriate control groups. This study aimed to mitigate this problem by using the following procedures: 1) using a bootstrapping approach for the biased clinical sample; 2) generating a random number dataset as a control population; 3) resampling both the bootstrapped targeted datasets and the normed control population; and 4) conducting a repeated analysis to create averaged statistics using the Monte Carlo simulation. The dataset used in the present study included 273 children with a history of delinquency and was assessed using the WISC-IV. Compared with conventional analyses, the proposed approach in the present study was found to generate the characteristics of the targeted clinical group on the basis of averaged statistics. Given that the norm had been identified in past research on psychometric intelligence, the use of bootstrapping and Monte Carlo simulations led to more robust findings compared with the use of conventional clinical studies.

## Keywords

Bootstrapping, Monte Carlo Simulation, Juvenile Delinquency, Intelligence Testing

---

## 1. Introduction

### 1.1. A Common Dilemma Faced by Clinical Psychologists

Despite the controversies associated with the Boulder model (e.g., Drabick &

Goldfried, 2000), clinical psychologists and education/training directors still generally use the scientist-practitioner model for their professional psychological activities (Norcross, Gallagher, & Prochaska, 1989; O'Sullivan & Quevillon, 1992). Psychologists practicing in the clinical field thus frequently share a common dilemma with regard to research methodology: sampling issues.

For example, clinical psychologists working in psychiatric hospitals routinely assess psychiatric patients using some psychological tests. Thus, they can accumulate the test data regarding psychiatric patients with comparative ease. However, the data they obtained have methodological shortcomings for scientific research.

Conventional survey designs are strongly recommended in psychological studies to obtain scientifically sound findings, both to collect a large, unbiased sample and to set control groups. However, clinical psychologists often face difficulties in assembling data for nonclinical participants (i.e., the control group) due to the clinical heterogeneity and small sample sizes of their routine casework. This inevitably means that in the absence of a large sample and a proper control group, the findings from such studies are not as scientifically robust as they could be.

## 1.2. A Prescription to Mitigate Sampling Issues

Simulation techniques have been used to solve sampling deficiencies in recent psychological research (e.g., Carpenter & Bithell, 2000; Rasmussen, 1989). One of the methodologies in computational statistics for addressing sampling issues is the use of random numbers, namely, simulation approaches (e.g., Del Moral, Doucet, & Jasra, 2012; Deng & Lin, 2000; Sitter, 1992). This study aimed to investigate the application of several computational simulation techniques that can hopefully contribute to clinical research findings, including bootstrap estimation and the Monte Carlo approach.

Bootstrapping is a resampling method that repeatedly uses a specific dataset (Efron & Tibshirani, 1986). Compared with studies where the collected dataset is only used once, the bootstrapping approach uses the dataset repeatedly in order to increase the reproducibility of the findings. Bootstrapping consists of the following procedures: 1) the research data is collected as part of the clinical study, and the obtained dataset is regarded as a population for the target group; 2) the data for each of the population is numbered in order; 3) another dataset is made by random sampling with replacement from the population, and this resampling process is repeated until the number of datasets is sufficient enough; and 4) averaged statistics are calculated within each of the datasets to provide parameter distributions of the target variables.

Interval estimates from resampling distributions are better than point estimates from an original dataset because they are generally composed of small and biased samples (Hall & Martin, 1988). The use of the bootstrap method can thus be particularly helpful to clinicians in a practical field limited to small clinical samples.

Another issue to contend with is that of control groups. It is difficult to set appropriate control groups in clinical studies because nonclinical people seldom visit clinical psychologists. One solution is the use of random number generation when the norm of the population is already known through previous standardization. If the norm statistics are equivalent to population parameters, a random number dataset can thus be simulated from the standardization sample. Each control group dataset could then be created from the generated random numbers of the population. In actual survey research, a control group does not always accurately reflect the true population. However, it is also inappropriate to use the simulated random number as a control group because the population dataset is generally too large and the targeted clinical data is too small. Instead, repeated random sampling from the total simulated population can be used to properly compare the two groups. The Monte Carlo approach (Doubilet, Begg, Weinstein, Braun, & McNeil, 1985) is a combination of the procedures above: the creation of infinite datasets through bootstrapping and random number generation to estimate and evaluate the true values of the given phenomena using the averaged statistics from repeated samplings.

This study aimed to determine the appropriate resampling times and the effect of the sample size using the Monte Carlo simulation on a sample of children with a history of juvenile delinquency. Findings from past studies regarding delinquent populations have shown a higher likelihood of deteriorated intelligence (e.g., McGloin, Pratt, & Maahs, 2004; Moffitt & Silva, 1988) and lower verbal abilities (e.g., Andrew, 1977; Isen, 2010). The purpose of this study was thus twofold: 1) to investigate the incremental efficacy of the Monte Carlo approach using bootstrapping compared with traditional statistical analyses and 2) to determine the appropriate procedures regarding the resampling times (Davidson & MacKinnon, 2000). The hypothesis of the present study was that low intelligence and lower verbal abilities in delinquent children could be replicated using the Monte Carlo simulation.

## 2. Methods

### 2.1. Procedure

The dataset of the relevant population was obtained from a Japanese child guidance center, a public institution where delinquent children under 14 years of age are referred to for clinical assessment and treatment. The prerequisite to be included in the study was intellectual ability as determined by the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) (Wechsler, 2010). A total of 273 children were included in the dataset.

The control group was created by using NtRand 3.3, an Excel add-in random number generator based on the Mersenne Twister algorithm (Numerical Technologies, 2017). NtRand 3.3 requires the mean and covariance of the objective variables in order to generate random numbers according to the multivariate normal distribution. The mean and covariance of the 10 subtests in the WISC-IV

Japanese version (Wechsler, 2010) were thus used to generate 100,000 random cases as a population. The computed scaled scores for the 10 subtests were adjusted as follows: if the calculated scaled score was less than 1 or over 19, the number was fixed to 1 and 19, respectively; the four indices, verbal comprehension index (VCI), perceptual reasoning index (PRI), working memory index (WMI), and processing speed index (PSI), were calculated as the sum of the 10 subtests according to the conversion table (Wechsler, 2010).

The clinical group was then compared with the control group as follows: the bootstrap method was applied to the clinical group to repeat the comparison. Random sampling with replacement for the 273 delinquent children was repeated several times: 10,000, 8000, 5000, 2000, 800, 500, 200, 80, 50, 20, 8, 5, and 2 times. For cross-validation purposes, the population size was operationally decreased to evaluate the differences from the results of the total data by 246 (90%), 218 (80%), 191 (70%), 164 (60%), 137 (50%), 109 (40%), 82 (30%), 55 (20%), and 27 (10%). For the control group, random sampling without replacement from 100,000 cases of the population was iterated to compare with the clinical group, and a same sample size was used as the clinical group.

Reiterated tests were finally conducted to compute the descriptive statistics ( $M$  and  $SD$  for VCI, PRI, WMI, and PSI in both groups) and the inferential statistics ( $F$ ,  $p$ ,  $\chi^2$  in MANOVA, Cohen's  $d$ , Hedges'  $g$ , for VCI, PRI, WMI, and PSI in  $t$ -tests). The statistics were calculated repeatedly and obtained as distributions ( $M$ ,  $SD$ , and 95%  $CI$ ).

The study was approved by the ethical review board, and given the retrospective design of the study, the need for written informed consent was waived.

## 2.2. Participants

The participants included in the study were children with a history of crime: 209 boys and 64 girls. The ages of the children ranged from 9 to 15 years old ( $M = 13.2$ ,  $SD = 1.4$ ). The cases of delinquency included the following: runaway (28), theft (77), violent incidents (46), sexual deviation (23), arson (25), theft of household money (8), bad companionship (8), drug addiction (2), truancy (2), and miscellaneous (13). Using the WISC-IV, the children's full-scale IQ ranged from 57 to 117 ( $M = 84.1$ ,  $SD = 11.6$ ). The descriptive statistics of the WISC-IV were as follows:  $M = 81.2$ ,  $SD = 12.2$  for VCI,  $M = 88.9$ ,  $SD = 13.3$  for PRI,  $M = 88.0$ ,  $SD = 13.1$  for WMI, and  $M = 91.8$ ,  $SD = 13.1$  for PSI.

## 2.3. Measurement

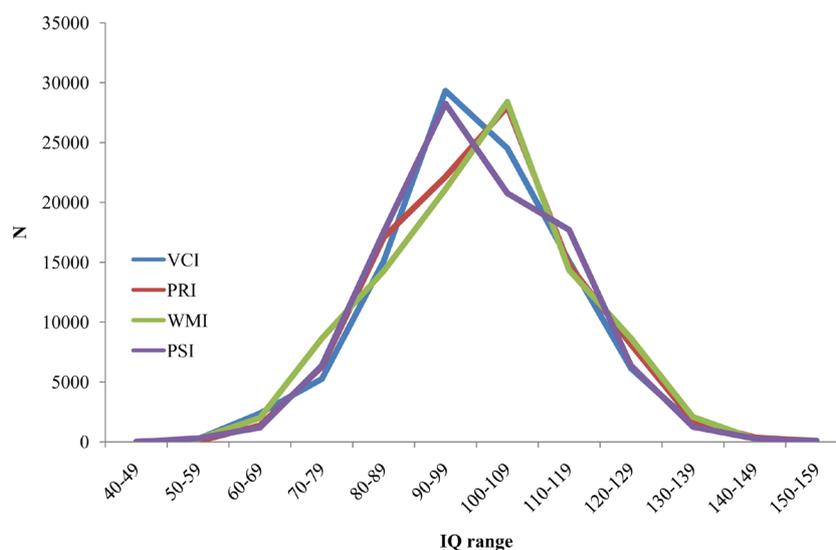
The Japanese version of the WISC-IV was standardized in 2010 based on the data of 1293 children (Wechsler, 2010). The model of four correlated factors was adopted theoretically to empirically substantiate the standardization study. The relationships between the 4 indices and 10 subtests were as follows: VCI, Similarities, Vocabulary, and Comprehension; PRI, Block Design, Picture Concepts, and Matrix Reasoning; WMI, Digit Span, and Letter-Number Sequencing; and PSI, Coding, and Symbol Search. The reliability coefficients based

on the split-half method were 0.90 for VCI, 0.89 for PRI, 0.91 for WMI, and 0.86 for PSI, and those based on the test-retest method ( $N = 88$ , interval  $M = 22$  days) were 0.91 for VCI, 0.78 for PRI, 0.82 for WMI, and 0.84 for PSI. The psychometric properties were considered adequate for the present study.

### 3. Results

#### 3.1. The Validity of the Control Population

The distributions and correlation matrices for the four indices were inspected and compared with those of the standardization population simulation in order to confirm the validity. **Figure 1** presents the approximate normality of distributions for VCI, PRI, WMI, and PSI. Given the large data size of 100,000, Kolmogorov–Smirnov tests to assess the normality of the data could not be performed; thus, the skewness and kurtosis of the four indices were used instead. **Table 1** shows that few differences were found from zero and the approximate equivalence between the present simulation and the norm regarding the correlation coefficients.



**Figure 1.** Distributions for VCI, PRI, WMI, and PSI by random number generation. VCI, verbal comprehension index; PRI, perceptual reasoning index; WMI, working memory index; PSI, processing speed index.

**Table 1.** Distribution properties and correlation matrices for VCI, PRI, WMI, and PSI by random number generation.

	VCI	PRI	WMI	PSI	Skewness	Kurtosis
VCI		0.49	0.46	0.29	0.05	0.09
PRI	0.49		0.47	0.34	0.12	-0.09
WMI	0.47	0.48		0.32	-0.04	-0.25
PSI	0.30	0.34	0.32		0.07	0.03

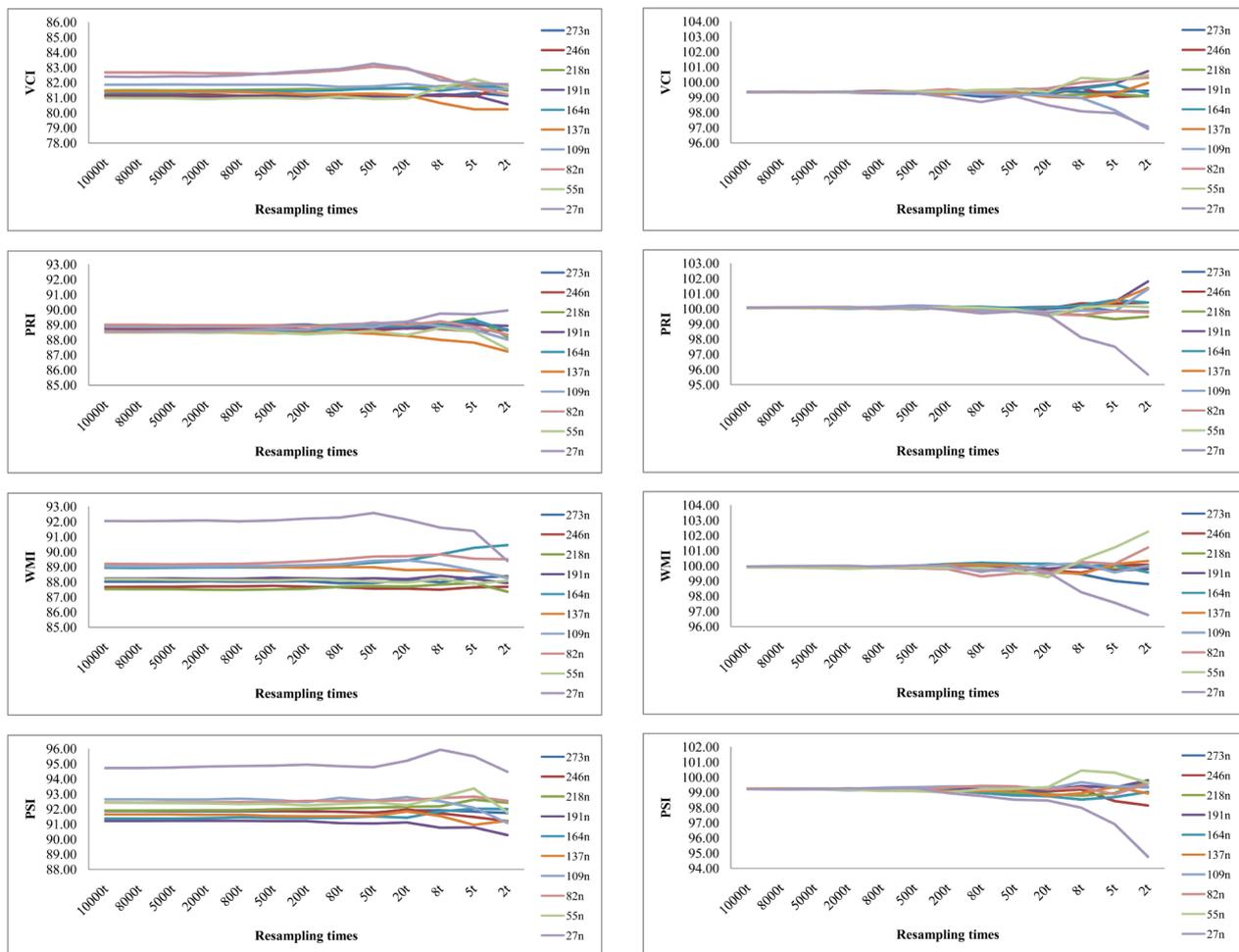
Note: The upper triangle indicates the results of the present simulation. The lower triangle indicates the results of the standardization study (Wechsler, 2010). VCI, verbal comprehension index; PRI, perceptual reasoning index; WMI, working memory index; PSI, processing speed index.

### 3.2. Simulation Results in Both Groups

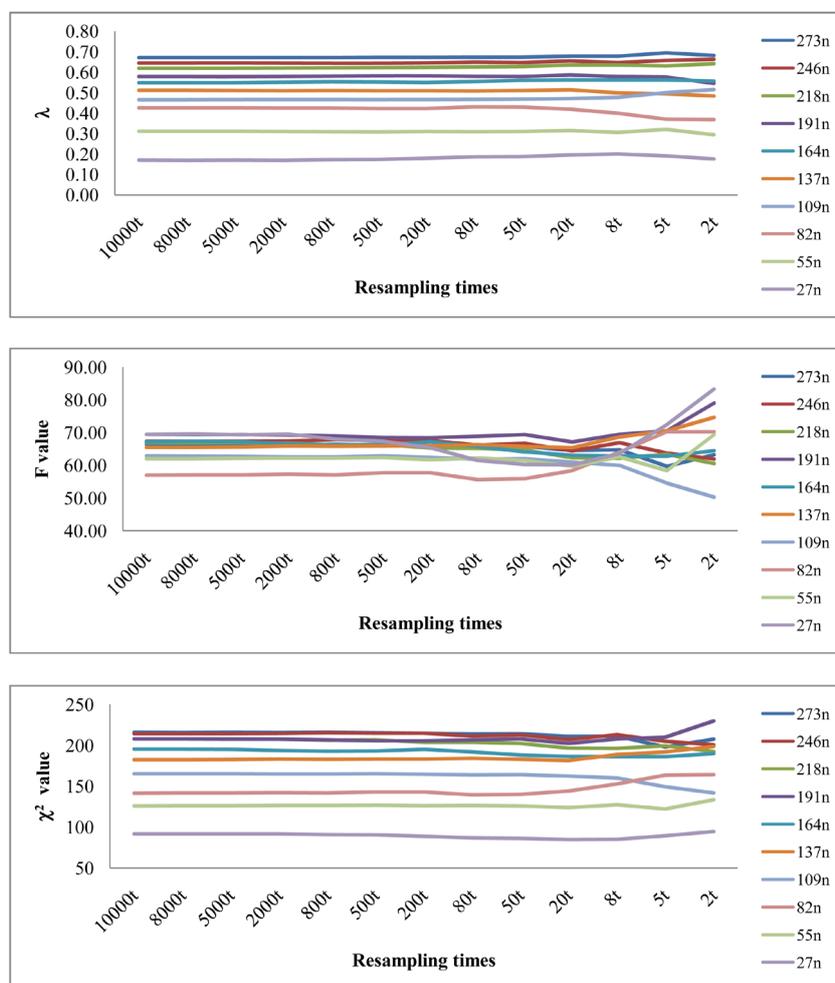
Concerning the four indices, **Figure 2** summarizes the mean variations determined by the sample size and the number of repetitions in both the clinical and control groups. Compared with the control group, the clinical group had a larger variance due to the sample size, and the particular accuracy of the mean estimates deteriorated when based on less than 70% of the total dataset (191n in **Figure 2**). On the other hand, the control group had more stable estimates when the sample size decreased. In order to make stable estimations for the clinical group, resampling had to be conducted more than 2000 times; anything less than 500 times made unstable estimations. For the control group, however, resampling more than 50 times was enough to make estimates stable (see **Figure 2**).

### 3.3. Differences between Groups

A multivariate analysis of variance (MANOVA) was employed to determine the overall differences between the two groups according to the four indices. **Figure 3**



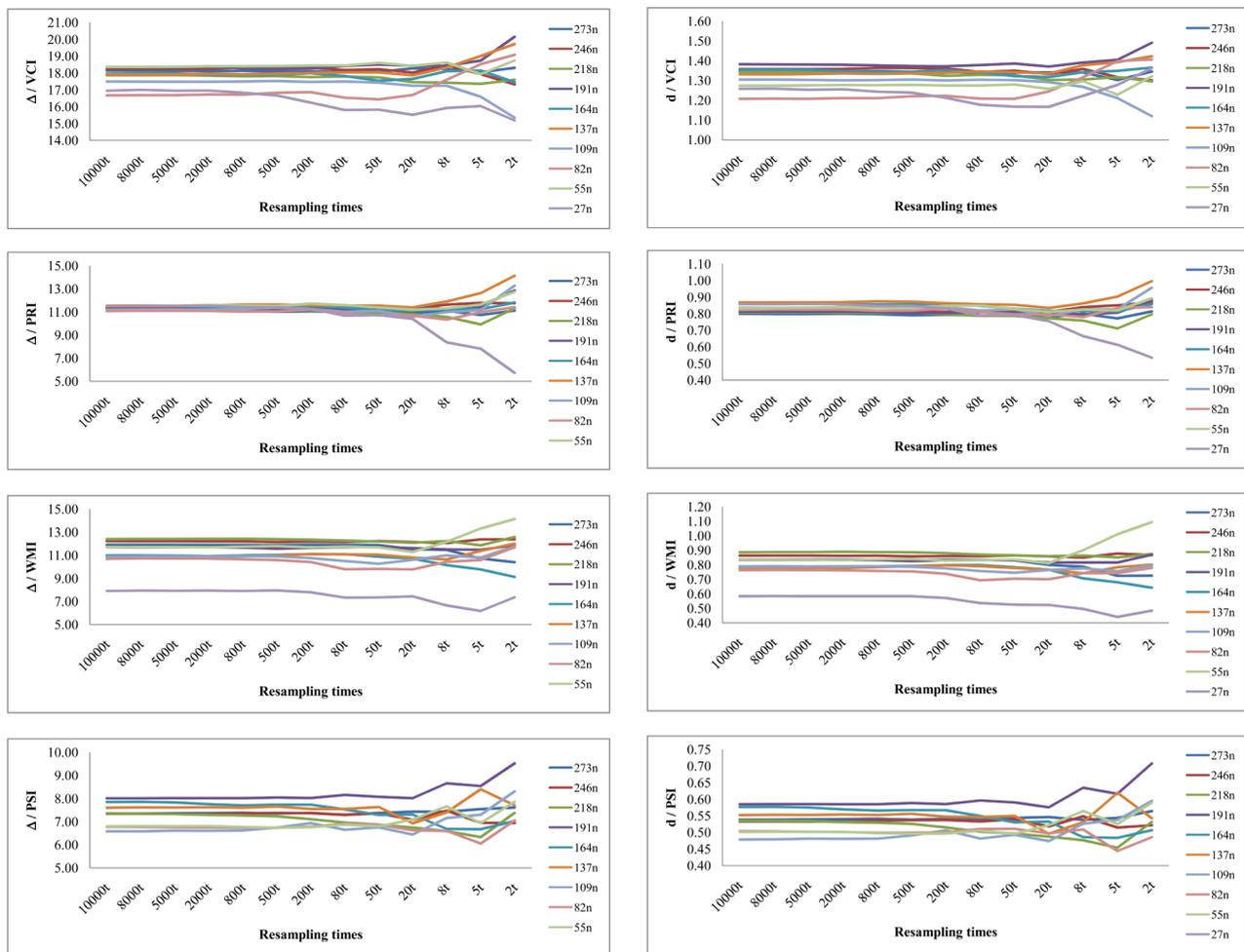
**Figure 2.** Mean variations of VCI, PRI, WMI, and PSI according to both sample size and repeated times. The clinical (delinquent) group is shown on the left-hand column and the control group on the right-hand column. VCI, verbal comprehension index; PRI, perceptual reasoning index; WMI, working memory index; PSI, processing speed index.



**Figure 3.** MANOVA statistics between clinical and control groups on VCI, PRI, WMI, and PSI. VCI, verbal comprehension index; PRI, perceptual reasoning index; WMI, working memory index; PSI, processing speed index.

presents the mean variation according to sample size and resampling times for Wilks' lambda ( $\lambda$ ),  $F$  value, and  $\chi^2$  value. With respect to  $\lambda$ , there were no variations according to resampling times, but there was a decreased effect according to sample size. With regard to  $F$ , the variance was larger when resampling was conducted less than 500 times, but the sample size had a relatively small effect. As far as the  $\chi^2$  value was concerned, less than 50% of the total sample size decreased the  $\chi^2$  value, and less than 200 resampling times made the stability of the mean statistics worse.  $P$  values for both  $F$  and  $\chi^2$  were less than 0.0000001 at least. The results indicated that there was a significant overall difference in the four indices between the two groups.

MANOVA was conventionally employed for individual profile analysis irrespective of statistical appropriateness (Bray & Maxwell, 1982; Enders, 2003; Warne, 2014). In the current study, profile analyses were conducted for the four indices separately (see Figure 4). The raw differences ( $\Delta$ ) were defined as the scores of the clinical group subtracted from those of the control group. The



**Figure 4.** Cohen's  $d$  differences between clinical and control groups for VCI, PRI, WMI, and PSI. VCI, verbal comprehension index; PRI, perceptual reasoning index; WMI, working memory index; PSI, processing speed index.

standardized differences (Cohen's  $d$ ) were defined as the mean differences between groups divided by the pooled  $SD$ .

For VCI, the simulation results were stable unless resampling was conducted less than 500 times or the sample size was less than 40% (109n). For PRI, the simulation results were stable unless resampling was conducted less than 50 times irrespective of the sample size. For WMI, the simulation results were stable unless resampling was conducted less than 200 times or the sample size was less than 10% (27n). For PSI, the simulation results were stable unless resampling was conducted less than 500 times or the sample size was less than 30% (82n).

### 3.4. Full Simulation Results

The results outlined above indicate that larger sample sizes and higher resampling times could improve the accuracy of the comparison using the Monte Carlo simulation. For this reason, the number of resampling times was set at 10,000 for the present study, and the full sample size (100%) was used. **Table 2** is a summary of the results of the Monte Carlo simulation:  $M$  and 95%  $CI$  for

**Table 2.** Comparative results using the Monte Carlo simulation.

	VCI	PRI	WMI	PSI
	95% L, M, 95% H	95% L, M, 95% H	95% L, M, 95% H	95% L, M, 95% H
Average				
Delinquent	79.7, 81.2, 82.6	87.4, 88.9, 90.5	86.5, 88.0, 89.6	90.3, 91.9, 93.4
Control	97.7, 99.3, 101.1	98.3, 100.1, 101.8	98.1, 99.9, 101.7	97.5, 99.2, 100.9
Comparison				
$\Delta$	15.9, 18.1, 20.4	8.8, 11.1, 13.5	9.5, 11.9, 14.3	5.1, 7.4, 9.6
Cohen's <i>d</i>	1.16, 1.35, 1.54	0.63, 0.80, 0.98	0.66, 0.84, 1.01	0.36, 0.54, 0.72
Hedges' <i>g</i>	1.16, 1.35, 1.54	0.63, 0.80, 0.98	0.66, 0.84, 1.01	0.36, 0.54, 0.72
<i>p</i>	0.00, 0.00, 0.00	0.00, 0.00, 0.00	0.00, 0.00, 0.00	0.00, 0.00, 0.00
MANOVA				
$\lambda$	0.61, 0.67, 0.73			
F	50.3, 66.4, 85.3			
<i>p</i>	0.00, 0.00, 0.00			
$\chi^2$	171.3, 216.1, 265.1			
<i>p</i>	0.00, 0.00, 0.00			

Note: 95% L, the lower limit of 95% CI; 95% H, the upper limit of 95% CI;  $\Delta$  mean difference between groups. All *p* values were at least less than 0.0001.

descriptive statistics (average), comparison statistics, and MANOVA. The 95% *CI* mentioned here denotes the 95th percentile: the lower limit of 95% *CI* was the 2.5th percentile, whereas the upper limit of 95% *CI* was the 97.5th percentile. The MANOVA statistics were all found to be significant, indicating that there were overall differences between the cognitive profiles of the two groups. The effect sizes of each of the indices demonstrated a small effect for PSI (0.36), a medium effect for PRI (0.63) and WMI (0.66), and a large effect for VCI (1.16).

### 3.5. Conventional Analysis

A conventional analysis was applied to compare the mean of the clinical group with the norm using a sample *t*-test with a constant for each of the four indices. The findings indicated that the results for the clinical group had significantly lower scores than the norm ( $M = 100$ ) for VCI ( $t [272] = 25.4$ ,  $p < 0.001$ , Cohen's  $d = 1.54$ ), for PRI ( $t [272] = 13.8$ ,  $p < 0.001$ , Cohen's  $d = 0.83$ ), for WMI ( $t [272] = 15.2$ ,  $p < 0.001$ , Cohen's  $d = 0.92$ ), and for PSI ( $t [272] = 10.3$ ,  $p < 0.001$ , Cohen's  $d = 0.62$ ).

## 4. Discussion

### 4.1. Appropriate Resampling Times

The goals of the present study include the following: to determine the number of resampling times required for stable statistical results and to investigate the de-

trimental effects of decreasing the sample size on the validity of the estimates. With regard to the former, a resampling of more than 2000 times was enough to reliably estimate the targeted statistics. Furthermore, given that the simulation results were resampled more than 2000 times, they were found to be appropriate for the Monte Carlo comparisons using bootstrapping procedures (Davidson & MacKinnon, 2000).

#### **4.2. Effects of Sample Size Reducing**

Reducing the sample size was also found to make unstable estimations after 70% or less of the total population was used. Given that all of the figures of any phenomena are never identifiable to compile a true dataset, it is necessary to thus complement sampling in case there are missing values for at least 30% of all participants.

However, given that virtual data collection methods are likely to be influenced by varying factors, including bias and clinical heterogeneity, more reiterations are needed to make the estimated statistics stable compared with control populations. The current simulation only recommends less than 2000 iterations as adequate due to the use of appropriate random sampling with smaller errors of measurement.

#### **4.3. Replication and Validity of the Demonstration**

The proposed methodology in the present study would not have been appropriate if a comparison of the simulation results did not detect lower IQ and verbal ability in the clinical group. However, the present study corroborated the findings of past research (Andrew, 1977; Isen, 2010; McGloin et al., 2004; Moffitt & Silva, 1988) (see **Table 2**). The overall intellectual ability of the children with a history of delinquency was lower than the norm because the 95% *CI* of all four broad abilities did not include the mean of 100. Furthermore, only VCI reached the borderline intellectual level on the basis of 95% *CI* (79.7, 82.6). Therefore, the actual survey did not deviate from the findings of past samples of children with a history of delinquency.

#### **4.4. Advantages of the Monte Carlo Simulation**

The advantages of using the Monte Carlo simulation in the present study were as follows: firstly, it yielded a distribution of statistics in the target group. Conventional research with a single sample can only compute a point estimation of the target clinical participants, and this inevitably requires researchers to swallow assumptions of the theoretical distribution to calculate confidence intervals. Furthermore, as mentioned previously, sampling in a clinical study tends to be frequently small and biased. Given that the Monte Carlo method can be used as part of clinical examinations, it can provide more robust statistics compared with point estimation. Secondly, multivariate analysis can be applied for sample investigations in routine clinical settings. Considering that an analysis without

simulation can only be compared with a standardized norm, the available statistical analyses are limited to simple comparisons to a given constant value (e.g., one-sample *t*-test). On the other hand, random number generation could allow clinicians to contrast the target group to a simulated control group using multivariate analyses. Finally, and most critical of all, the present study demonstrated that the simulation strategy was as valid as the prior examination, in which an actual survey was carried out: that is, the results replicated the representative findings with regard to intelligence testing of the delinquent group.

Due to the above findings, it is strongly recommended that clinical psychologists consider the use of the simulation method for their research in order to increase the robustness of their findings using the bootstrapping and Monte Carlo simulations.

#### 4.5. Social and Practical Suggestions

The current findings suggest that clinical psychologists in practical fields should not abandon their research works due to the difficulty in sampling issues. Using the Monte Carlo methods on the basis of the present findings, they can analyze their routine practices scientifically and study the research theme they have interest in irrespective of sampling difficulties.

Consequently, the findings have possibilities to promote the scientist-practitioner model in clinical psychologist education. Adopting the computational statistics technique as a new methodology may expand the scientific expertise for clinical psychologists.

#### 4.6. Limitations and Future Research

Some defects in the present methodology must be noted. Firstly, this method cannot be used in clinical studies unless the norm is previously known and standardized scales are available. Bootstrapping procedures also have limitations in estimating the true values of a theoretical population. Although bootstrapped distributions may have more validity for the target clinical population relative to one sampling result, the results of the analyses can be influenced by the given fundamental sampling. Although not a perfect solution, the bootstrapping is a relatively robust methodology to navigate the sampling issues for research conducted in clinical settings.

Furthermore, in considering the proposed strategies relating to sampling issues and the simulation applied in this study, it is also necessary to consider the limitations of resampling in frequentist statistics. Obtained resampling data are usually independent each other; and thus any compensations have not done irrespective of repeating times. Given that the Bayesian approach can correct and update the probabilities along with the increasing number of estimation times (e.g., [Alfaro, Zoller, & Lutzoni, 2003](#); [Smith & Gelfand, 1992](#)), the prior probabilities can be theoretically near to the true values. Future research is thus desirable to compare the present methodology with the Bayesian approach in the context of clinical studies.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- Alfaro, M. E., Zoller, S., & Lutzoni, F. (2003). Bayes or Bootstrap? A Simulation Study Comparing the Performance of Bayesian Markov Chain Monte Carlo Sampling and Bootstrapping in Assessing Phylogenetic Confidence. *Molecular Biology and Evolution*, *20*, 255-266. <https://doi.org/10.1093/molbev/msg028>
- Andrew, J. M. (1977). Delinquency: Intellectual Imbalance? *Correctional Psychologist*, *4*, 99-104. <https://doi.org/10.1177%2F009385487700400108>
- Bray, J. H., & Maxwell, S. E. (1982). Analyzing and Interpreting Significant MANOVAs. *Review of Educational Research*, *52*, 340-367. <https://doi.org/10.3102/00346543052003340>
- Carpenter, J., & Bithell, J. (2000). Bootstrap Confidence Intervals: When, Which, What? A Practical Guide for Medical Statisticians. *Statistics in Medicine*, *19*, 1141-1164. [https://doi.org/10.1002/\(SICI\)1097-0258\(20000515\)19:9%3C1141::AID-SIM479%3E3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-0258(20000515)19:9%3C1141::AID-SIM479%3E3.0.CO;2-F)
- Davidson, R., & MacKinnon, J. G. (2000). Bootstrap Tests: How Many Bootstraps? *Econometric Reviews*, *19*, 55-68. <https://doi.org/10.1080/07474930008800459>
- Del Moral, P., Doucet, A., & Jasra, A. (2012). On Adaptive Resampling Strategies for Sequential Monte Carlo Methods. *Bernoulli*, *18*, 252-278. <https://doi.org/10.3150/10-BEJ335>
- Deng, L. Y., & Lin, D. K. (2000). Random Number Generation for the New Century. *The American Statistician*, *54*, 145-150. <https://doi.org/10.1080/00031305.2000.10474528>
- Doubilet, P., Begg, C. B., Weinstein, M. C., Braun, P., & McNeil, B. J. (1985). Probabilistic Sensitivity Analysis Using Monte Carlo Simulation: A Practical Approach. *Medical Decision Making*, *5*, 157-177. <https://doi.org/10.1177/0272989X8500500205>
- Drabick, D. A., & Goldfried, M. R. (2000). Training the Scientist-Practitioner for the 21st Century: Putting the Bloom Back on the Rose. *Journal of Clinical Psychology*, *56*, 327-340. [https://doi.org/10.1002/\(SICI\)1097-4679\(200003\)56:3%3C327::AID-JCLP9%3E3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4679(200003)56:3%3C327::AID-JCLP9%3E3.0.CO;2-Y)
- Efron, B., & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, *1*, 77. <https://doi.org/10.1214/ss/1177013817>
- Enders, C. K. (2003). Performing Multivariate Group Comparisons Following a Statistically Significant MANOVA. (Methods, Plainly Speaking). *Measurement and Evaluation in Counseling and Development*, *36*, 40-56. <https://doi.org/10.1080/07481756.2003.12069079>
- Hall, P., & Martin, M. A. (1988). On Bootstrap Resampling and Iteration. *Biometrika*, *75*, 661-671. <https://doi.org/10.1093/biomet/75.4.661>
- Isen, J. (2010). A Meta-Analytic Assessment of Wechsler's P>V Sign in Antisocial Populations. *Clinical Psychology Review*, *30*, 423-435. <https://doi.org/10.1016/j.cpr.2010.02.003>
- McGloin, J. M., Pratt, T. C., & Maahs, J. (2004). Rethinking the IQ-Delinquency Relationship: A Longitudinal Analysis of Multiple Theoretical Models. *Justice Quarterly*,

- 21, 603-635. <https://doi.org/10.1080/07418820400095921>
- Moffitt, T. E., & Silva, P. A. (1988). IQ and Delinquency: A Direct Test of the Differential Detection Hypothesis. *Journal of Abnormal Psychology, 97*, 330-333. <https://doi.org/10.1037/0021-843X.97.3.330>
- Norcross, J. C., Gallagher, K. M., & Prochaska, J. O. (1989). The Boulder and/or the Vail Model: Training Preferences of Clinical Psychologists. *Journal of Clinical Psychology, 45*, 822-828. [https://doi.org/10.1002/1097-4679\(198909\)45:5%3C822::AID-JCLP2270450521%3E3.0.CO;2-E](https://doi.org/10.1002/1097-4679(198909)45:5%3C822::AID-JCLP2270450521%3E3.0.CO;2-E)
- Numerical Technologies (2017, October 19). *NTRAND 3.3: An Excel Add-In Random Generator Powered by Mersenne Twister Algorithm*. Word Press. <http://www.ntrand.com/>
- O'Sullivan, J. J., & Quevillon, R. P. (1992). 40 Years Later: Is the Boulder Model Still Alive? *American Psychologist, 47*, 67-70. <https://doi.org/10.1037/0003-066X.47.1.67>
- Rasmussen, J. L. (1989). Computer-Intensive Correlational Analysis: Bootstrap and Approximate Randomization Techniques. *British Journal of Mathematical and Statistical Psychology, 42*, 103-111. <https://doi.org/10.1111/j.2044-8317.1989.tb01118.x>
- Sitter, R. R. (1992). A Resampling Procedure for Complex Survey Data. *Journal of the American Statistical Association, 87*, 755-765. <https://doi.org/10.1080/01621459.1992.10475277>
- Smith, A. F., & Gelfand, A. E. (1992). Bayesian Statistics without Tears: A Sampling-Resampling Perspective. *The American Statistician, 46*, 84-88. <https://doi.org/10.1080/00031305.1992.10475856>
- Warne, R. T. (2014). A Primer on Multivariate Analysis of Variance (MANOVA) for Behavioral Scientists. *Practical Assessment, Research & Evaluation, 19*, Article No. 17. <http://pareonline.net/getvn.asp?v=19&n=17>  
[https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1329&context=pa\\_re](https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1329&context=pa_re)
- Wechsler, D. (2010). *Technical and Interpretive Manual for the Wechsler Intelligence Scale for Children* (4th ed.). K. Ueno, K. Fujita, H. Maekawa, T. Ishikuma, H. Dairoku, & O. Matsuda, Trans., Nihon Bunka Kagakusha.