DOCUMENT RESUME

ED 335 394 TM 017 010

AUTHOR	Lai, Morris K.; Saka, Thomas
TITLE	A Post-Hoc Procedure That Can Be Better than Random Assignment to Treatment.
PUB DATE	Apr 91
NOTE	15p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Comparative Testing; *Control Groups; *Experimental Groups; *Followup Studies; Grade 3; *Kindergarten Children; Primary Education; *Research Methodology; *Sampling
IDENTIFIERS	Hawaii; Missouri Kindergarten Inventory Develop Skills; Peabody Picture Vocabulary Test (Revised); *Post Hoc Methods; Randomization; Stanford Achievement Tests

ABSTRACT

An analytical procedure that uses fall kindergarten assessment data to retroactively create equivalent comparison groups for longitudinal research and evaluation studies was designed and tested, and then analytically and empirically compared with random assignment to treatment. Students entering kindergarten in Hawaii are administered the Peabody Picture Vocabulary Test (Revised) and the Missouri Kindergarten Inventory of Developmental Skills. In 1990, 8,909 matched pairs of students (i.e., students with 1986 fall kindergarten data and 1990 third-grade data) with data on these tests and on the Stanford Achievement Test in grade 3 were identified, and retroactive equivalent groups were created to compare a group who had received some sort of specific treatment with equivalents. Given the distribution of pretreatment scores for the treatment group, along with socioeconomic data and ethnic distribution, an equivalent distribution was created by filling the slots with appropriate individuals, also choosing the schools so as to ensure equivalence in distribution of school attrition rates. The net result was two groups with equivalent distributions. Had the design relied on random assignment to treatment, there would have been non-trivial probabilities that the groups would differ substantively on one or more of the variables. The method appeared feasible for conducting reality-based research when data allow for selection of a comparison group retroactively. Three tables present study data. (SLD)

A Post-hoc Procedure that can be Better than Random Assignment to Treatment

Morris K. Lai, University of Hawai'i
Thomas Saka, Hawai'i State Department of Education
5 April 1991

Paper presented at the annual meeting of the American Educational Research Association, Chicago.

A Post-hoc Procedure that can be Better than Random Assignment to Treatment

Morris K. Lai, University of Hawai'i
Thomas Saka, Hawai'i State Department of Education

Objectives

The objectives of this study were to (1) design and test an analytical procedure that uses fall, kindergarten assessment data available statewide to retroactively create equivalent comparison groups for longitudinal research and evaluation studies and (2) analytically and empirically compare the procedure with random assignment to treatment.

Background and Perspective

Many previous attempts at using matching to compensate for nonrandom assignment have been reported as not successful (Willing, 1985). Campbell and Stanley (1963) asserted that educational researchers should reject "the concept of achieving equation through matching (as intuitively appealing and misleading as that is)" [p. 2]. Since that recommendation was made, many other conflicting assertions have appeared in the professional literature.

Here we present a chronological rendition of selected quotations from the literature on matching and randomization. This is then followed by a summary of the quotations.

Quotations from the Literature on Matching and Randomization

"Perhaps Fisher's most fundamental contribution has been the concept of achieving pre-experimental equation of groups through randomization. This concept, and with it the rejection of the concept of achieving equation through matching (as intuitively appealing and misleading as that is) has been difficult for educational researchers to accept." [Campbell & Stanley (1963), p. 2]

"An absence of randomization may in some specific way plausibly explain the obtained results. But unless one can specify such a hypothesis and the direction of its effects, it should not be regarded as invalidating." [Campbell & Ross (1970), p. 123].

"There is no excuse for not including randomization where it is feasible to do so. The purpose [of the authors' article] has been to emphasize that randomization is not infallible, that fallibility increases to the extent that the probability of repeated experiments is low, and that this low probability of repeated experiments is particularly likely in the case of large-scale social interventions." [Sherwood, Morris, & Sherwood (1975), p. 220]

"...if the matching is done with presumed balance and carried out on an intuitive basis by the researcher, the principle of random allocation must be viewed as violated, and the effect on subsequent findings will be unknown. In general, there is very little to support the principle of attempting to achieve balanced groups on the basis of either intuition or expert opinion. The reason for this is that the biases in the allocations are not known, and, intrinsically, the procedure carried out is likely to be one that is highly complex." [Borgatta (1979), p. 164]

"Within a certain probability level, pretest equivalence can be achieved through randomization. However, it should still be determined if group equivalency following assignment has been ascertained." [Grinnell (1981), p. 579-80]

"Essentially, inappropriate applications occur when matching is carried out on the basis of premeasures of the outcome variables used to assess impact...However, matching on the basis of other variables is feasible and desirable (Sherwood et al., 1975)." [Rossi & Freeman (1982), p. 222]

"Even though matching alone does not adequately establish equivalence between groups, this procedure is still inappropriately used in the literature. A simpler procedure to increase the equivalence between groups is to carry out matching along with random assignment, or to conduct an analysis of covariance using pertinent variables as covariates." [Moore (1983), p. 173]

"Randomization only starts experimental groups out in the right way...There are plenty of opportunities for important confounding variables to creep into the design while the study is being conducted...One hopes that as confounding variables are identified (usually in retrospect), their probable effects can be judged either by common sense or from the research literature and in that way be taken into account." [Porter (1988), p. 401]

"Matching procedures usually create more problems than they solve. You cannot be certain that you have selected the most important variable or variables on which to match subjects. Also you may not be able to find suitable matches for some members of the characteristic-present sample. Therefore, the preferred procedure is to try to select the characteristic-present and comparison samples randomly from the same population, and then to control for other variables through the use of analysis of covariance..." [Borg & Gall (1989), p. 545]

"Perhaps our intuition better "matches" reality than we have been giving it credit for, at least in the realm of research design, matching, and random assignment to treatment." [Lai & Saka (1991), p. 8]

Who, When	**What They Said** [***comments by Lai & Saka***]
Campbell & Stanley, 1966	Don't use matching (even though intuitively appealing). [***Given as dogma more than with justification***]
Campbell & Ross, 1970	Lack of randomization is not necessarily invalidating. [***Matching no longer always forbidden***]
Sherwood et al., 1975	Randomization is likely to be fallible for large-scale interventions. [***Gave a complicated matching procedure***]
Borgatta, 1979	Intuition or expert opinion for matching is worthless. [***Experts we know are not that inept***]
Grinnell, 1981	Even after randomization, should check for equivalency. [***Only accept if the groups match?!***]
Rossi & Freeman, 1982	Matching OK if not on premeasure of outcome variable. [***They cite regression to mean; but isn't it controlled?***]
Moore, 1983	Use matching & random assignment or ANCOVA. [***ANCOVA is no panacea (assumptions?)***]
Porter, 1988	Use common sense or research literature to judge probable effects of confounding variables despite randomization. [***Still implies requirement of randomization, but definitely more respectful of experts***]
Borg & Gall, 1989	Preferred to matching is randomization with ANCOVA. [***Maybe they read Moore's book?***]

Matching as empirically and intuitively appealing

We assert that matching may be empirically as well as intuitively appealing. If the treatment being studied is not inextricably confounded with geographic location, then the use of pre-treatment scores available for the population, ethnicity (the importance of including this variable has been documented by Richman and Millar (1984) and Roth (1984)), and socioeconomic status (SES) can be actually an improvement over the use of random (even stratified) assignment to treatments.

Indeed Peterson, DeGracie, and Ayabe (1987, p. 109) not only successfully used a matched comparison group in a longitudinal study on retention, but they also chose not to delete cases pairwise inasmuch as they "could think of no compelling reason why any of the matching variables or membership in either group [retained or promoted] would differentially influence attrition."

In the state of Hawai'i, which has just one school district, approximately 98% of eligible-age children are enrolled in kindergarten. As part of the state's early childhood education program, all students are assessed upon entering kindergarten with the *Peabody Picture Vocabulary Test, Revised (PPVT-R)* and the following subtests of the *Missouri Kindergarten Inventory of Developmental Skills (MKIDS):* Number Concepts, Auditory Skills, Paper/Pencil Skills, Language Concepts, Visual Skills, and Gross Motor Skills. During the 1986-87 school year when a state norming study was conducted, the entering kindergartners' language pretest mean corresponded to a student at about the 19th percentile, while the posttest mean at the end of kindergarten corresponded to a student at about the 38th percentile.

In contrast to their normatively low performance at the beginning of kindergarten, Hawai'i's third graders statewide have performed close to the 50th percentile on the *Stanford Achievement Test* (SAT). It therefore follows that many of the students who had scored well below national norms upon entering kindergarten must have ended up scoring above national norms at the third grade, albeit on a different standardized test.

In this paper we use the *PPVT-R* and *SAT* to help us address the following question: Is it possible to improve upon random assignment to treatment by retroactively creating equivalent comparison groups?

Methods/Data source

MKIDS and *PPVT-R* data were available for approximately 14,600 children who were entering kindergarten during the fall of 1986. In the spring of 1990, about 13,000 third-grade students took the Reading, Vocabulary and Mathematics subtests of the *Stanford Achievement Test* (7th edition). We used students' 10-digit identification number (assigned by the Hawai'i Department of Education) to match the data from the two testing periods. All told, 8,909 matched pairs (i.e., students with 1986 fall, kindergarten data and 1990 third-grade data) were found. An additional 4,337 students with 1986 kindergarten test data did not have 1990 third-grade data, and 4,469 students with 1990 third-grade data did not have 1986 kindergarten data.

Means in terms of the various universes of students are shown in the following table.

Insert Table 1 about here

The exact method used to create equivalent comparison groups would depend on the type of comparison being made. For example, if a researcher wanted to study a cooperative mastery learning approach that was used by all teachers in, say five elementary schools, then he/she would find five other schools with similar fall, kindergarten test score distributions and similar socioeconomic status (including ethnicity) distributions but which had not used cooperative mastery learning.

On the other hand, if within the same school, "effective" kindergarten teachers have been identified (e.g., on the basis of their students' showing substantially larger than average gains on the pre-post kindergarten tests) then it may be possible to (retroactively) locate from the same school an equivalent (e.g., same SES and pretest distribution) group of students whose kindergarten teachers had not been exceptionally "effective" as defined previously. In order to address reliability concerns, the study might be designed to use two years' worth of data to identify "effective" teachers as well comparison group teachers.

In this paper we investigate the degree to which it is possible to create various types of comparison groups to address research/evaluation questions such as listed earlier. We expect that there will be substantial variance in the difficulty of creating different types of comparison groups.

Results

Analytic Arguments

The use of population-wide fall, kindergarten assessment data to retroactively create equivalent groups can, for a number of reasons, be an improvement over the usually impossible-to-obtain random assignment to treatment. First, the procedure presented in this paper controls for the following internal sources of invalidity cited in Campbell and Stanley's (1963) classic treatise: history, maturation, testing, instrumentation, regression, selection, mortality, and interaction of selection and maturation, etc. Other obstacles in conducting randomized experiments in field setting have discussed by Cook and Campbell (1979).

Second, it can be argued that the design also controls for the external sources of invalidity such as the reactive or interaction effect of testing. In effect the fall, kindergarten testing is part of the treatment, and thus there is no "unpretested universe from which the experimental respondents were selected." *Just as is the case for many true experimental designs using random assignment to treatment,* the design being discussed in this paper may not fully control for (a) the interaction of selection and treatments whose effects are being studied, (b) the reactive effects of experimental arrangements, which would preclude generalization about the effect of the experimental variable upon persons being exposed to it in nonexperimental settings, and (c) the multiple-treatment interference that may occur because the effects of prior treatments are not usually erasable.

Note that even if random assignment to treatment had been possible in the example of a cooperative mastery learning approach being used by all teachers in five elementary schools, there could have been unsurmountable problems such as if some of the "control" teachers started using cooperative mastery learning techniques (e.g., through adaptation of a program that unexpectedly included major aspects of cooperative mastery learning). The post-hoc method being presented in this paper, however, would not only guarantee equivalence on variables such as entering-kindergarten assessment data and SES but also guarantee that the comparison group would not have used a program with a strong cooperative mastery learning bent to .

In the second example, random assignment of students to kindergarten teachers within the same school would not work if the "control" teachers changed grade level after one year. The proposed post-hoc procedures could, however, "demand" (retroactively) that comparison teachers with two years worth of kindergarten teaching be found.

Empirical evidence of problems with random sampling

We took entire populations of students and computed mean NCE scores of 34.5 (*PPVT-R* in kindergarten), 48.1 (*Stanford Achievement Test*, Grade 3 Reading), and 54.0 (*Stanford Achievement Test*, Grade 3 Mathematics). We then selected straight random and stratified (by the seven subdistricts) random samples of various sizes (50 to 1000) and computed corresponding means, standard deviations, and *t*-tests.

Sampling of students.

Relatively large discrepancies occurred when we compared the *PPVT-R* means of the various sampling types. The rationale here is that looking at worse-case scenarios is warranted because if it is not that unlikely that large differences in estimating the (true) population mean can occur, then it is correspondingly not that unlikely that researchers who use random sampling would be dealing with "randomly equivalent" samples that in actuality differed substantially.

With the aforementioned rationale in mind, we compared the means of the seven types of sampling conducted. In the worse-case scenario, the kindergarten means differed as much as 4.1 NCE points (N=100 for each sample). This difference corresponded to almost a quarter of a standard deviation or about 2.5 standard error (of the mean) units.

Next we generated ten different random samples of 100 students taken from the population of those with kindergarten pretest data. In essence we retroactively created a random assignment to treatment setup wherein the "treatments" were whatever happened to the two groups during the four years between the start of kindergarten and the end of grade 3. In the absence of evidence to the contrary, we must assume that any one group of randomly selected students did not receive a more "effective" treatment than did any other group.

As shown in Table 2, among the ten samples, the largest difference in pretest means was 8.3 NCEs, which represents about .4 of a standard deviation or about 4 standard error (of the mean) units. Of the ten sample means, three were larger than 36.67, and five were smaller than 33.67. Thus several presumed equivalent means would have been more than three NCEs, and, therefore, more than a standard error (of the mean) apart.

Insert Table 2 about here

Perhaps more important was the amount of "equivalence" that was lost over the four years. Attrition in the ten randomly equivalent samples ranged from 23% to 38%.

Now we turn to these students' third-grade performance on the *Stanford Achievement Test* (*SAT*). Again the assumption is that no group had a particularly more effective treatment than any other group. What we find in fact is that the mean *SAT* mathematics scores ranged from 49.7 NCEs to 56.8 NCEs, and the mean reading scores ranged from 46.0 NCEs to 50.0 NCEs.

We now turn from random assignment to random sampling. From all students in the population, we randomly sampled ten groups of 100 each. The number of students with pretest scores ranged from 62 to 79, with corresponding mean NCE pretests of 27.8 and 35.5 respectively. Mean third-grade mathematics NCE scores ranged from 51.0 to 62.0; reading NCE means from 46.6 to 52.2. Findings are summarized in Table 3.

Insert Table 3 about here

Sampling of schools.

School mean differences of a few NCEs can be more critical (statistically) because of the relative narrowness of the distribution of the means of groups. The difference between the mean NCE of a random sample of 50 schools and the mean of the population was about a fourth of a (group) standard deviation or about 1.75 standard error (of the mean) units. The worse-case scenario for the mathematics means occurred in the comparison between the random sample of 5 schools vs. the random sample of 10 schools. In that case, the difference in mean NCEs was 4.5, which corresponds to about a third of a standard deviation.

Attrition with regard to individual students can be viewed as a dichotomy: A 1986 kindergartner with fall *PPVT-R* data either does or does not have 1990 *SAT* data. Schools, however, can be viewed as having an attrition rate theoretically between 0% and 100%. If schools are randomly assigned to treatment, differential attrition rates could lead to substantial non-equivalence in later grades.

We calculated two types of attrition rates for schools: (1) of the students with fall, 1986 kindergarten *PPVT-R* data, the percent not having spring, 1990 third-grade data on the *SAT*; and (2) of the students with spring, 1990 third-grade *SAT* data, the percent not having fall, 1986 kindergarten *PPVT-R* data. The category (1) attrition rates for schools ranged from 21% to 98%, while the category (2) rates ranged from 19% to 99%. Some of the extreme rates are relatively easily explained (e.g., large transient military population or isolated areas with relatively little movement into the area or out of the area); however, rates at many sites are not readily predictable. Perhaps, a new school was built nearby or a nearby industry was shut down.

The point being made here is that if schools are randomly assigned to a treatment, there is a good chance that there will be differential attrition rates that cannot always readily be anticipated. In our data set four schools are close enough geographically to have the same name stem. Yet these four schools had category (2) attrition rates ranging from 41% to 70%.

Stratified random sampling can be an improvement, but...

For the school sampling, we had socioeconomic data available in the form of percent of students whose families received Aid to Families with Dependent Children and General Assistance. When this variable was used to stratify the sampling, the resulting mean NCEs were closer to the population mean than had been the case for the straight random sampling; however, the corresponding mean mathematics NCE was farther away from the population mean than was the case for the straight random sampling of 25 schools.

It is somewhat ironic that stratified random sampling is well accepted as an improvement over straight random sampling because matching is much like stratification carried out to an extreme. In essence stratification is designed to help ensure better matching when random sampling is being conducted.

A Post-hoc Procedure that can be Better than Random Assignment

The procedure delineated in this paper has been proven successful in retroactively creating equivalent groups. It is given here in a generic form with the full realization that each use would have to be customized somewhat.

Suppose we have a group of students who have received some sort of specific treatment, and we wish to find an equivalent group for comparison purposes. Under the circumstances we have outlined in this paper, pre-treatment data are available for the population of students. Given the distribution of pre-treatment scores for the treatment group, along with the socioeconomic status and/or ethnic distribution of these students (all on a data tape), an equivalent distribution is created by filling in slots with the appropriate individuals. If one-to-one matching on a variable is not possible, then categories such as high, middle, low (e.g., SES) may be used. The schools from which the students are picked will be chosen so as to ensure equivalence in the distribution of school attrition rates.

The net result is having two groups whose distributions are equivalent on, for example, distribution of *PPVT-R* (or some other population-available pre-treatment test) scores upon entering kindergarten, SES, attrition rate at school, and ethnicity. As we have shown, if the design had relied on random assignment to treatment, there were non-trivial probabilities that the groups would differ substantively on one or more of the aforementioned variables.

Educational importance and concluding remarks

We have shown that it is feasible to retroactively create equivalent groups tor use in comparing effects of some schooling variables. Furthermore there are empirical evidence and analytical arguments that the procedure can be an improvement over random assignment to treatment. Matching matches reality and may be better than randomization when used retroactively.

Using longitudinal data (kindergarten to grade three), we found noteworthy instances of non-equivalence when randomization (random assignment or random sampling) was used or retroactively "created." Some variables such as attrition rate cannot be adequately predicted longitudinally, and thus randomization could easily lead to non-equivalence on those variables, whereas the proposed post-hoc procedure could guarantee longitudinal equivalence on identified variables. There is substantial expertise about what variables to (retroactively) match on. In short we have a method amenable for use in conducting reality-based research in which comparison groups are made as equivalent as possible.

As with other research methods, there are limitations in the procedure discussed in this paper. It will not be possible to always find an appropriately equivalent comparison group; however, if one can be found, then it may turn out that using such a group may enable us to do better research than even if we had been able to obtain random assignment to treatment. Perhaps our intuition better "matches" reality than some have been giving it credit for, at least in the realm of research design, matching, and random assignment to treatment.

References

Borg, W. R., & Gall, M. D. (1989). *Educational research.* New York: Longman.

Borgatta, E. F. (1979). Methodical considerations–Experimental and non-experimental designs and causal inference. In T. Abramson, C. K. Tittle, & L. Cohen (Eds.). (1979). *Handbook of vocational education evaluation* (pp. 161-178). Beverly Hills: Sage.

Campbell, D. T., & Ross, H. L. (1970). The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. In E. R. Tufte (Ed.), *The quantitative analysis of social problems*. Reading, MA: Addison-Wesley.

Campbell, D. T., & Stanley J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation*. Chicago: Rand McNally.

Grinnell, R. M. (1981). *Social work research and evaluation.* Itasca, IL: F. E. Peacock Publishers.

Lai, M. K., & Saka, T. (1991, April). A post-hoc procedure that can be better than random assignment to treatment. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Moore, G. W. (1983). *Developing and evaluating educational research.* Boston: Little, Brown, and Company.

Petersen, S. E., DeGracie, J. S., & Ayabe, C. R. (1987). A longitudinal study of the effects of retention/promotion on academic achievement. *American Educational Research Journal, 24*(1), 107-118.

Porter, A. C. (1988). Comparative experiments in educational research. In R. M. Jaeger (Ed.), *Complementary methods for research in education*, 391-411. Washington, DC: American Educational Research Association.

Richman, B. O., & Millar, G. W. (1984). What I think and feel: A cross cultural study of anxiety in children. *Psychology in the Schools, 21*, 255-57.

Rossi, P. H., & Freeman, H. E. (1982). *Evaluation–A systematic approach*. Beverly Hills: Sage.

Roth, R. (1984). Schooling, literacy acquisition and cultural transmission. *Journal of Education, 166*, 291-308.

Sherwood, C. C., Morris, J. N., & Sherwood, S. (1975). A multivariate, nonrandomized matching technique for studying the impact of social interventions. In E. L. Struening & M. Guttentag (Eds.), *Handbook of evaluation research, Volume 1* (pp. 183-224). Beverly Hills: Sage.

Willing, A. E. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research, 55*, 269-317.

Table 1. NCE means and standard deviations for various universes of students

	N	PPVT-R Mean (S.D.)	SAT Reading Mean (S.D.)	SAT Mathematics Mean (S.D.)
K & Gr. 3 data	8,909	35.1 (22.8)	48.8 (18.9)	55.0 (21.8)
K data only	4,337	33.5 (22.8)		
Gr. 3 data only	4,469		48.0 (19.3)	53.2 (21.3)

Table 2. Random Samples (N=100) of Students with *PPVT-R* Scores

	PPVT Mean	(S.D.)	*SAT Reading* Mean	(S.D.)	N	*SAT Mathematics* Mean	(S.D.)	N
Random Sample 1	33.3	(22.5)	48.0	(17.9)	62	54.6	(20.0)	63
Random Sample 2	37.0	(23.6)	50.2	(18.5)	72	55.8	(20.1)	72
Random Sample 3	33.9	(23.1)	46.0	(19.9)	70	49.7	(22.5)	70
Random Sample 4	36.7	(22.0)	50.0	(18.0)	74	56.3	(19.3)	73
Random Sample 5	32.4	(23.2)	47.7	(16.4)	77	55.5	(20.1)	77
Random Sample 6	33.7	(23.5)	49.4	(16.6)	63	56.8	(16.7)	62
Random Sample 7	33.8	(21.3)	47.0	(15.2)	67	51.2	(20.2)	66
Random Sample 8	28.6	(21.1)	47.9	(20.4)	63	54.1	(22.5)	63
Random Sample 9	36.9	(22.2)	48.2	(15.6)	66	54.2	(22.0)	66
Random Sample 10	33.4	(23.4)	49.0	(19.7)	72	56.2	(22.7)	72

Table 3. Random Samples (N=100) of Students in the Population

	PPVT Mean	(S.D.)	N	*SAT Reading* Mean	(S.D.)	N	*SAT Mathematics* Mean	(S.D.)	N
Random Sample 1	35.5	(22.2)	71	47.6	(20.4)	73	51.8	(19.5)	73
Random Sample 2	33.2	(23.2)	71	50.5	(18.9)	72	55.7	(23.6)	72
Random Sample 3	27.8	(21.1)	62	49.1	(18.4)	81	55.5	(20.2)	81
Random Sample 4	35.5	(24.3)	79	50.0	(18.4)	75	56.2	(22.1)	75
Random Sample 5	33.0	(22.7)	69	47.8	(20.5)	78	51.0.	(21.8)	78
Random Sample 6	34.7	(22.1)	63	52.2	(19.7)	76	62.0	(23.0)	76
Random Sample 7	32.9	(20.3)	64	48.9	(18.3)	75	54.3	(18.7)	75
Random Sample 8	34.7	(22.4)	69	48.2	(17.9)	78	53.1	(21.6)	78
Random Sample 9	33.9	(22.4)	71	49.0	(20.0)	63	56.9	(21.0)	63
Random Sample 10	36.0	(24.8)	64	46.6	(17.7)	77	51.0	(22.7)	77