THE 2000 HOUSING UNIT DUPLICATION OPERATIONS AND THEIR EFFECT ON THE ACCURACY OF THE POPULATION COUNT

Robert E. Fay[1]
U.S. Census Bureau, 4700 Silver Hill Rd. Stop 9001, Washington, DC 20233-9001

KEY WORDS: Census, Census Undercount, Census Adjustment, Record Linkage

1. Introduction.

Major aspects of the implementation of Census 2000 reflected years of planning and preparation. But a number of elements were revised following the 1999 Supreme Court decision that statistical adjustment could not be incorporated into the census counts used for apportionment. Unrelated to the court decision, however, the Housing Unit Duplication Operations (called here the *duplicate operation*), was one of the most improvised components. The duplicate operation was designed and implemented to address a potentially serious problem in the preliminary results from Census 2000. Not only was the operation the last major step to be designed, but it also lacked any close precedent in previous censuses. The operation also had a substantial effect on the census: It subtracted approximately 3.6 million from the early census population count, although most of the removed records corresponded to persons included elsewhere in the census. The Census Bureau is currently conducting evaluations of the operation.

The paper will address two different but interrelated perspectives on the analysis of the duplicate operation:

(a) evaluating the duplicate operation itself, to assess the degree to which it succeeded in improving the accuracy of the census counts; and
(b) using results from the duplicate operation to investigate the Accuracy and Coverage Evaluation (A.C.E.), the sample survey designed to estimate the net error of population counts from Census 2000.

Both perspectives are concerned generally with the accuracy of the census counts, but the perspectives differ on the implications of counting individual persons more than once.

With respect to (a), decisions on design of Census 2000 and previous censuses have rested in part on the *housing unit model*. The housing unit model also guided the design of the duplicate operation. Under the housing unit model, the first objective of the census is a complete and accurate inventory of all housing units and other places where people live. (The census also includes procedures for group quarters and the homeless.) Determination of the number of occupants in each housing unit and their personal characteristics are also important objectives, but they must rest on an accurate housing count. The paper will describe the manner in which the housing unit model influenced the design of the duplicate operation. Specifically, the housing unit model was the rationale for including in the census counts some apparently duplicated individuals when the associated housing units appeared distinct.

With respect to (b), the A.C.E. was designed to measure person coverage in Census 2000 from an independent sample of persons and a sample of census enumerations. Matching the sample of persons to the census on the basis of individual reported characteristics was a critical aspect of the A.C.E. design. The coverage of housing units was not a central component of the A.C.E.'s approach to estimating the net error of the population count. (But a parallel study of the accuracy of the housing unit count was imbedded in the A.C.E.) In the A.C.E., duplicate enumeration of persons in the census is counted as an error without regard to the housing unit model.

The paper details how the A.C.E. sample offers opportunities to assess aspects of the duplicate operation. It also describes how the computer matching methods used for the operation could be redirected to assess the A.C.E. These opportunities include the use of matching 1) to assess the quality of some components of respondents' reports in A.C.E. interviews, 2) to examine the uniformity of the effects of duplication over the poststrata used for A.C.E. estimation, and 3) to account for the effect of possible universe differences between the A.C.E. and the census.

The duplicate operation used two types of computer matching: exact matching on first and last name and on month, day, and year of birth; and a statistically-based matching with the Fellegi-Sunter algorithm as implemented by the Statistical Research Division at the Census Bureau. Exact matching was used to identify potential duplicates over wide areas, while the statistically-based matching was used to evaluate the similarity of households matched either by the sharing of one or more exactly matching persons or by the similarity of addresses. The paper analyzes data from the operation to evaluate the implications of this strategy. In particular, the analysis indicates that statistically-based matching may find approximately 60% more duplicated households than exact matching. Nonetheless, statistically-based matching has the potential for yielding substantially more incorrect matches than exact matching if it is applied too widely, especially to the entire country. This paper is intended to summarize information on duplication available from the operation itself. Other researchers are

currently conducting expanded matching efforts to further investigate these issues.

2. Duplication in Previous Censuses

There are two methods to estimate the undercount of the U.S. censuses, demographic analysis and survey-based methods, such as the 2000 A.C.E. Demographic analysis (Robinson et al. 1993) uses demographic relationships and data from past censuses, registered births and deaths, and other administrative sources to estimate the population of the U.S. for each census year. Although there have been some efforts to develop methods for subnational estimates, demographic analysis has been primarily recognized for its account of the national population.

The A.C.E. is the most recent effort to measure the undercount through sample survey methods. The 2000 A.C.E. shares several features in common with the 1990 Post Enumeration Survey (PES); these two surveys incorporated significant methodological advances over predecessors in 1950 and 1980. The surveys generally attempted to estimate net census error by measuring census omissions and erroneous enumerations at the level of individual persons. Thus, the survey approach provides an indication of the components of gross error in the census. (The surveys have been designed for the measurement of net error, and they have not attempted to measure gross error consistently from one census to the next.) In general, the estimated components of omissions and erroneous enumerations have been larger than the estimated net undercount.

Duplication represents a significant component of erroneous enumerations. The duplication of housing units and of persons frequently occurs together, but each form of duplication also occurs separately. For example, the same persons may be enumerated in Apt. 1 and Apt. 2, leading to duplication of persons but not housing units. Similarly, duplicate housing units may be enumerated once as occupied and once as vacant, leading to duplication of housing units but not persons.

Comparison of measured duplication rates for recent censuses provides some perspective on the results for 2000, but the estimates are not entirely comparable because of methodological variations in the coverage measurement surveys. In 1980 the census was organized into enumeration districts (EDs), each an area, typically of several blocks, representing an expected followup workload for one enumerator. The 1980 Post Enumeration Program (PEP) estimated a person duplication rate of approximately 0.8% within ED or in adjacent EDs in the designated area of search (Cowan and Fay 1984). Additionally, an estimated 1.0% of 1980 enumerations were classified as *geocoding errors,* when the census placed them outside their correct ED and also outside the PEP area of search. A subsequent study estimated that approximately 30% of this group, or 0.3% of the overall census count, duplicated enumerations in the correct ED. Hence, the 1980 PEP indicated a total rate of duplication of approximately 1.1% in the same ED or geographically nearby EDs. The overall estimate of erroneous enumerations in 1980 was 3.4%.

In the 1990 PES, the area of search was the census block and adjacent census blocks. In urban areas, a single ring of adjacent blocks constituted the search area, whereas in rural areas two rings did. Consequently, the 1990 search area is not comparable to 1980, although 1990 areas should generally have tended to be smaller than 1980 (*1*). The 1990 PES estimated a duplication rate within the search area of approximately 1.6% (Hogan 1993). (In addition, the 1990 imputation method estimated 0.3% additional erroneous enumerations. Because the imputation method did not further assign a type of erroneous enumeration, some proportion of the 0.3%, perhaps 0.1%, implicitly represented imputed duplicate enumerations.) The geocoding error rate was estimated to be 0.3%, lower than in 1980. The overall estimate of erroneous enumerations in 1990 was 5.8%.

In 1990, a related study, the Housing Unit Coverage Study, estimated a duplication rate for housing units of 1.0% (Childers 1993b). The estimated overall erroneous enumeration rate was considerably higher for vacant units, 9.2%, than occupied, 2.2%, but the study did not report a distribution of duplicates by occupied/occupied, occupied/vacant, and vacant/vacant pairings, so it is not clear how many of the housing unit duplicates were a probable cause of person duplication, which would occur only with occupied/occupied pairings.

The A.C.E. definition of duplicate enumeration was similar to 1990, but there were two notable differences. First, although all sample clusters were searched for duplicates within the cluster, searches for duplicates in surrounding blocks were conducted only for a sample of clusters in the A.C.E., unlike 1990. Second, in 1990, two rings of surrounding blocks were searched in rural areas, but in 2000, only a single ring was. A single ring of blocks was searched in urban areas in both 1990 and 2000.

Because of the effect of the duplicate operation, the estimate of duplicates from the A.C.E., 0.8%, is not comparable to 1990. The concluding section accounts for this lack of comparability.

3. Duplication of Housing Units in Early 2000 Master Address Registers

Evaluation results from the two previous censuses thus suggested that some duplication was unavoidable. In 1999 and 2000, however, evidence accumulated for a potentially high level of duplication in Census 2000. Nash (2000) later reviewed the decisions and context of the development of the Master Address File (MAF) for

Census 2000, emphasizing considerable differences from 1990. A strategy of multiple sources for addresses, combined with conservative rules for eliminating potential duplicates, sought to reduce omissions of housing units at the risk of potential duplication. Nash remarked, "In doing this, the Census Bureau assumed responsibility for developing a comprehensive, unduplicated file of addresses."

Evidence of housing unit duplication accumulated from comparison of counts from the MAF to independent housing unit estimates. Site visits in June 2000 to selected areas, including New York City, Baltimore, and Chicago, provided specific examples. Census 2000 was also the first to capture name and date of birth in a computer readable form; census staff (Donald Dalzell and David Word) detected a surprisingly large number of exact matches within a sample of census enumerations.

As Nash (2000, p. 3) noted, Nonresponse Followup, which was still in the field, had the potential to eliminate some duplicates, but would likely overlook others. For example, households receiving two questionnaires and returning both by mail would probably not be eliminated by the operation. Because of the multiple modes of response--mail, internet, personal visit, telephone, and Be Counted Forms--the census design already incorporated a Primary Selection Algorithm (PSA) to resolve duplicate enumerations of persons linked to the same MAF identifiers (ids). But the PSA only operated within ids and was not designed to detect or eliminate duplicated addresses in the MAF.

4. Design of the Housing Unit Duplication Operations

The Census Bureau designed and implemented the duplicate operation during the summer and fall of 2000 to correct a potential overcount of housing units, and consequently of persons, in Census 2000. Two primary methods were used to identify potential housing unit duplications.

- Address matching based on characteristics of the unit addresses recorded in the MAF. After a period of testing, 7 edit rules (*2)* were selected (Knott and Galdi 2000). (For example, Apt 1A and #A1 at the same address were linked.) A method of scoring the goodness of match between any pair of addresses was developed and used in later steps.
- Person matching based on name and date of birth. Matches were only considered within state and excluded the group quarters population. Housing units paired by one or more exact person matches were then evaluated for the strength of match between other persons in the two households. Vacant units were excluded from person matching. Later, a method of evaluating the strength of person match between any two paired households was developed.

A.C.E. requirements influenced the design of the duplicate operation. A.C.E. timing required the definition of an A.C.E. universe by October 2000, even if this universe did not include everyone in the final census count. A two-phase operation accommodated the timing requirements. Phase 1 was to take an inclusive approach to identifying potential duplicates, including all pairs that were likely candidate duplicates. For each pair, triple, etc. of candidate duplicates, one housing unit was selected to remain, and others were provisionally deleted. The A.C.E. universe excluded these provisional deletions, and no attempt was made to match A.C.E. sample persons to them. Phase 2 was then allowed to reinstate any provisionally deleted units that did not appear duplicated after review. The dual-system estimator in the A.C.E. was modified to account for the omission of these reinstated units from the A.C.E. universe.

- In Phase 1, address matching and person matching were conducted independently. (Address matching was not performed in Puerto Rico.) The address matching using the 7 edit rules yielded 1,573,606 pairs. Person matching used exact matching followed by rules to evaluate the similarity of remaining persons within the potentially matching households. After excluding dissimilar households, person matching yielded 2,088,197 pairs. The two sets of results were combined into a file of 3,333,285 pairs (including Puerto Rico, Nash 2000, pp. 4-6). Most duplication resulted in single pairs, but triplets, quadruplicates, etc., were represented by multiple pairs.

  For each pair, triplet, etc., of housing units remaining in the census after deletions from field operations, including Nonresponse Followup, one was selected for inclusion in the census. The selection was by a deterministic algorithm, generally on the basis of a larger population count or more completely reported characteristics. The selected unit was guaranteed a place in the final file and included in the A.C.E. universe. The other household or households were marked as provisional deletions and excluded from the A.C.E. universe.
- In Phase 2, completed in November 2000, additional information on address matching and person matching was combined to decide which provisionally deleted units to reinstate. Phase 2 was also implemented virtually entirely as a computer operation *(3)*.

  The separate findings of address and person matching were assembled for each pair, along with second scorings of the address and person matching. The second scoring for address matching used rules that identified some of the addresses paired by person matching as similar (Knott and Galdi 2000). The second person matching used a modified version of Census Bureau matching software to identify some households as matching on a person basis, even though

Phase 1 had not found an exact match on name and birthday for any of the members (Fay 2001), or had previously excluded the pair as too dissimilar. Thus, pairs linked by either person matching only or address matching only in Phase 1 could be classified as matching on both a person and address basis in Phase 2.

The reinstatement rules for Phase 2 used the combined evidence from address and person matching from both phases, along with additional characteristics from the MAF and census files, to determine whether units provisionally deleted by Phase 1 should be reinstated. Except for a few global rules, separate rules were specified for the mailout/mailback census and those for other types of enumeration areas (*4)*. As noted in the Introduction, the rules reflected an overarching principle that may be described as the *housing unit model*. Under the housing unit model, duplicate enumeration of persons under some circumstances is regarded as an inevitable aspect of census taking. For example, if census forms are misdelivered to opposite units in a duplex and only one of the two households responds by mail, a census enumerator may accidentally obtain a second interview from the respondents while leaving the nonresponding household out of the census. Childers (1993a) used the 1990 Housing Unit Coverage Study to estimate that approximately 23% of census omissions according to the definitions of the 1990 PES were in omitted housing units, showing that omission of housing units was an important component of omission of persons. (The estimate excluded persons moving between Census Day and the initial 1990 PES interview.) The reinstatement rules generally reinstated households with duplicated persons in which it appeared that two separate units indeed existed. On the other hand, if the evidence suggested that the housing unit had also been duplicated, then the reinstatement rules permanently deleted the duplicate unit and any associated person records from the census.

Phase 2 reinstated 1,002,951 housing units and permanently deleted 1,371,320 from the U.S. count. (In Puerto Rico, 16,106 were reinstated and 11,366 deleted.) It reinstated 2,315,553 persons and permanently deleted 3,572,799 from the U.S. (In Puerto Rico, 50,587 were reinstated and 71,171 deleted.)

5. Assessing the Effect of the Housing Unit Duplication Operations

The Housing Unit Duplication Operations came about for two reasons without precedent in previous censuses: The development of the MAF for Census 2000 appeared to accept too many addresses from multiple sources without a sufficient strategy for eliminating duplicates, and the capture of name and date of birth for the first time permitted computer matching to help rectify the problem. It is unlikely that the operation will be built into the 2010 census in a similar manner--the Census Bureau plans develop a design to avoid the problem at earlier stages.

Nonetheless, the duplicate operation was an important component of Census 2000, and a number of related evaluations are currently underway to evaluate its impact. Among the issues are the following:

- Did the duplicate operation remove approximately the right level of housing units overall?
- Because the Phase 2 rules were quite different for the mailout/mailback census compared to other areas, did the methods achieve the correct balance between the mailout/mailback areas and others? (Mailout/mailback areas were identified by city-style addresses such as 101 Main St., while units in other areas were identified in relation to maps ("map-spots").)
- Were areas with high concentrations of Phase 1 duplicates treated appropriately relative to areas with low levels, so that after Phase 2 the resulting census counts had a more uniform level of coverage than before?

Some studies will focus specifically on the effect on the housing unit count, without regard to population coverage. A limited amount of relevant data may be available from the Housing Unit Coverage Study, which is based on the same sample as the A.C.E.

Although the duplicate operation was designed under the housing unit model, evaluation of the success of the operation in performing its primary purpose, to eliminate duplicate housing units, will not completely address the effect on the population count. For example, if the census enumerates the same household in two adjacent housing units, the one in which they live and a neighboring vacant unit, then the population count would be more accurate if the unit was deleted from the census, but the housing unit count is more accurate if the unit is retained.

6. Relationship to the A.C.E.

Besides duplication of persons in the same area, a different form of duplication occurs when a person is enumerated at a second address outside of the area of search. For example, persons who move near Census Day may be reported at both their old and new addresses. In 1980, 1990, and 2000, the coverage studies did not identify this form of duplication separately, but the interviews of the sample of census enumerations attempted to determine whether the persons were correctly enumerated at the sample address. The presumption is that when persons were duplicated, one of the two addresses was incorrect and the enumeration of the persons at the incorrect address would be identified as erroneous when the address was sampled. This approach did not separately confirm that the persons were duplicated, however. Thus, a full account of the effect of

duplication on the census count has not been available.

A project is underway to match the A.C.E. sample to the entire census. Evidence from this project could provide evidence on several aspects of the A.C.E.

- To assess the quality of some components of respondents' reports in A.C.E. interviews. Coverage measurement surveys, including the A.C.E., have generally required that respondents accurately report their residence on Census Day. In the 1990 PES, a reinterview months after the initial interview provided the primary evidence on the accuracy of the original reports, but the accuracy of the reinterview itself is easily questioned. Person matching will identify pairs of census enumerations, some of which will be included in the A.C.E. sample. Accurate reporting in the A.C.E. requires that essentially ½ of these pairs be reported as erroneous enumerations. Furthermore, the accuracy of the 2000 reinterview can also be investigated by the same logic.
- If the duplicate operation is found to have extracted too many or too few records from the national total or locally, then variation in the match rates or some other measure may be indicative of unevenness in census coverage. The A.C.E. was designed to adjust the census by applying a proportional adjustment within each of a set of poststrata. It is possible that detailed results from the operation will provide a symptomatic indication of the degree of variation within poststrata.
- Unexpected differences between the A.C.E. estimate for the total population and demographic analysis may stem in part from universe differences. For example, if the A.C.E. identifies students who live in college dormitories and are reported on their parents' census forms, their enumeration at home should be counted as erroneous. The accuracy of reports may be assessed by examining the A.C.E. reports for students duplicated in the census at college. The same questions extend to other duplications occurring between the A.C.E. and group quarters populations in the census.
- A component of the A.C.E., the Targeted Extended Search, may have missed some duplicates that should have been reflected in the estimation. The computer matching may identify the extent of this effect.

7. Limitations of Exact Matching

Exact matching limits although does not eliminate false matches. But exact matching potentially misses many of actual duplicate enumerations. Data from the duplicate operation offer some evidence on the extent to which exact matching may understate the actual level of person duplication.

In Phase 1, the 7 geographic edits paired addresses in the MAF thought to indicate potential duplication. Person match results for addresses linked by one of the 7 geographic edits are reported in Table 1 *(5)*. The results confirm that 5 of the 7 (edits 1, 2, 4, 5, and 6) successfully identified probable housing unit duplicates in mailout/mailback census areas, because each yielded duplicated households at rates above 50%. None of the edits did as well in areas outside of the mailout/mailback areas.

Across most of the edits, exact matching appears to identify only roughly 60% of the total matches indicated by the statistically-based matching (Table 1). Although the statistical matching employed an empirically derived cutoff score to separate matches from nonmatches, the distribution of scores was highly bimodal for this population, indicating relatively sharp discrimination between matches and nonmatches. (The cutoff score was 4. Overall, 41.1% scored below 0 as clear nonmatches, and 51.4% scored above 8 as clear matches.)

If Phase 1 required an exact match to link households, then a heuristic argument suggests that large households would be brought together more successfully than small ones, simply because there were more opportunities to match. Because Phase 1 and Phase 2 incorporated different penalties for matching households of different sizes, comparison of match results for linked households of the same size provides the most direct test of the heuristic argument. In Table 2 *(6)*, single-person households are the least likely to be recognized by exact match, at a rate slightly below 50% of all matches, while the rate for households of 3 or more is above 70%. The results for different size households suggests that the statistical-match employed in Phase 2 was far more accepting of matches between different size households than the algorithm used in Phase 1.

8. Discussion

The Housing Unit Duplication Operations substantially affected the population count from Census 2000, and currently a number of evaluations investigate their effect. Some of the findings from the A.C.E. will help to analyze the effect of the Housing Unit Duplication Operations. Person matching now offers some opportunity to investigate assumptions underlying the A.C.E.

Data from the duplication operation suggests that exact matching understates the level of actual person duplication. Data from the A.C.E. should provide additional evidence on this question, because duplicates identified clerically in the A.C.E. can be compared to results of computer matching in the same search area.

As previously noted, the two-phase approach to the duplicate operation accommodated timing requirements of the A.C.E. All units provisionally deleted in Phase 1 were excluded from the A.C.E. universe. Reinstated units from Phase 2 were added back into the census but not the A.C.E. universe. Consequently, A.C.E. estimates of duplication omit duplication to reinstated units.

Note 1: This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

The author would like to thank Philip Steel and William Winkler for their comments.

Notes

(*1*) The 1990 PES considered adjacent blocks without regard to ED, whereas in 1980 adjacent blocks in different EDs were often not in the same search area. Thus, the 1980 estimate of 0.8% duplication within the area of search may omit some cases counted in 1990. The estimate of 1.1% duplicates in 1980, which includes allowance for duplicates outside the search area, should include these. Thus, 1980 and 1990 results are not readily comparable, but the results indicate a higher level of duplication in 1990.

(*2)* Edits (Knott and Galdi 2000): (1) Potential MAF duplicate file (paired by a previous geographic match, but not deleted earlier from concern over possible loss of valid housing units); (2) Same Block, Address, and Within Structure ID, Different ZIP; (3) Same Block, House Number, and Within-Structure ID, Similar Street Name; (4) Same TIGER/Line ID and Side, Number, and Within-Structure Descriptor; (5) Within Structure Equivocation #1A, Apt A1; (6) House Number Suffix Problems 101A Main vs 101 Main Apt A; (14) Equivocation of within structure ids, for example, "A" = "1" or "First", etc.

(3) Practically speaking, Phase 1 could only have been a computer operation, because it involved all of the addresses and enumerations in the census. But for Phase 2, a clerical approach was possible and considered, because Phase 1 narrowed the number of pairs to examine. The clerical approach was tested but abandoned in favor of a second computer operation, except that headquarters staff resolved a small number of cases that involved a high number of pairings.

(*4)* Addresses in mailout/mailback areas were almost exclusively city-style (street number/street name) and those in other areas relied on map spotting and other enumeration. Consequently, address matching performed quite differently in the two types of areas.

(5) The full version of Table 1 is available from the author.

(6) Table 2 is available from the author.

References

Childers, D. (1993a), "The Impact of Housing Unit Coverage on Person Coverage, HUCS Results Memorandum Number 2," U.S. Census Bureau memorandum from Ruth Ann Killion to Thomas C. Walsh, June 24, 1993.

_____ (1993b), "Coverage of Housing in the 1990 Decennial Census, 1990 Decennial Census Preliminary Research and Evaluation Memorandum No. 253," U.S. Census Bureau, Oct. 7, 1993.

Cowan, C.D. and Fay, R.E. (1984) "Estimates of Undercount in the 1980 Census," *1984 Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington, DC, pp. 566-571.

Fay, R. (2001), "Method for Matching Persons During Phase 2 of the Duplicate Housing Unit Operations," draft U.S. Census Bureau memorandum, Jan. 16, 2001.

Hogan, H. (1993), "The 1990 Post-Enumeration Survey: Operations and Results," *J. of the Amer. Statist. Assoc.,* **88**, 1047-1060.

Hogan, H. (2000), "Specification for Reinstating Addresses Flagged as Duplicates on the Hundred percent Census Unedited File (HCUF)," DSSD Census 2000 Procedures and Operations Memorandum Series #D-11, U.S. Census Bureau, Nov. 7, 2000.

Knott, J. and Galdi, D. (2000), "Geography Division (GEO) Participation in Census 2000 Overcoverage Research: Processing of Address- and Person-Matched Files, U. S. Census Bureau memorandum, November 8, 2000.

Nash, F.F. (2000), Overview of the Duplicate Housing Unit Operations," unpublished document, Nov. 7, 2000.

Robinson, J. G., Ahmed, B., Das Gupta, P., and Woodrow, K. A. (1993), "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis, *J. of Amer. Statist . Assoc.,* **88**, 1061-1071.

U.S. Census Bureau (2001), "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy," March 1, 2001, available at http://www.census.gov/dmd/www/pdf/Escap2.pdf

Table 1. Person-match status of occupied housing units linked by geographic edits in Phase 1 of the housing duplicate operation, showing the match rate and the proportion of the total identified by exact matching.

Person Match Status of Pair	Geographic Edit Number						
	1	2	3	4	5	6	14
% matched/occ. pairs	76%	76%	36%	74%	73%	75%	31%
% exact/matched	63%	63%	64%	64%	54%	60%	47%