

Evaluating Merge/Purge Systems: Part Six

*By Jim Wheaton and Cynthia Baughan Wheaton
Principals, Wheaton Group*

Original version of an article appeared in the December 1987 issue of "Direct Magazine"

[Note: Despite dramatic increases in raw computing power and a proliferation of end-user software tools since the publication of this series of six articles, virtually all of the content remains highly relevant. The occasional obsolete point is highlighted.]

Statement of

In a series of six articles, we explain a number of the key concepts that mailers should understand about merge/purge, as well as reviewing (in the first article) a methodology that could be helpful in evaluating the effectiveness of either present or prospective merge/purge systems. While our comments are primarily addressed to mailers, merge/purge vendors can benefit by measuring themselves against the criteria that we have identified as important.

Our objective is to describe new and specific tools that can be used to evaluate and improve the performance of the merge/purge process. Through commentary and examples, we will attempt to translate into layman's terms the technical jargon that baffles many mailers. In the process, practical applications should become apparent.

This month's article, Part Six, focuses on the four methods of handling outside list duplicates, as well as suggested ways of monitoring the performance of duplicates that are mailed multiple times.

Methods of Handling Outside List Duplicates

The merge/purge process divides the gross names input into three primary groups:

- Duplicates to house or other suppression lists.
- Rented list duplicates, either between multiple lists (inter-list duplicates) or within a single list (intra-list duplicates).
- Unique net names available to mail.

The disposition of suppression list hits is easy. They should not be mailed. Rented list duplicates are not so straightforward. They can be handled in one of four basic ways:

- Random allocation.
- Hierarchical allocation.
- Random allocation, within a hierarchy.
- Separation of duplicates.

Although all four will be discussed, the third method, random/hierarchy, is recommended for most established direct mail companies.

Random Allocation

Here, duplicates are allocated on a random basis among the specific lists that contributed them.

- If, for example, List A had a 1,000-name overlap with List B, each list would be credited with 500 of these duplicates¹.
- If List C, D and E had 300 names in common, each would receive credit for 100 names.

Random allocation is particularly appropriate in a start-up mode, where all lists are untested and there is no existing reason to assign priority.

Exhibit 1 illustrates the impact on response reports of random allocation.

Hierarchical Allocation

In this case, lists are prioritized for duplicate allocation. Whenever a name overlap appears between two or more lists, the highest-ranked list is credited with the duplicate. Assume, for example, a hierarchy of A, B, C, D, E:

- List A would receive credit for all names that overlap with – say – A, B and C.
- List C would be credited all names that appear on – say – C, D and E.

The frequent motivation behind the hierarchical allocation of duplicates to outside lists is the minimization of list rental charges.

- Where, for example, List A has a rental charge of \$50 per thousand and List B a fee of \$100 per thousand, and each is rented on a net basis, it is economical for List A to receive the entire overlap.
- Many list owners charge on a gross name basis for first-time tests of small quantities (i.e., 5,000 to 10,000). Where, for example, List C is rented on a gross name basis and List D on a net name basis, and the quoted cost per thousand is identical, total charges are minimized when all duplicates are allocated to List C.

¹ For purposes of simplification, names are assumed to appear on no other lists but the ones cited in the examples and attachments.

The problem with this approach is that there is no reason intrinsic to the nature of the lists for a hierarchical setup. The by-product is distorted response reports and incorrect rollout decisions, which can more than offset the savings in list rental charges.

Exhibit 1 illustrates the impact on response reports of hierarchical allocation.

Random Allocation Within A Hierarchy

This approach, a permutation of the first two, combines multiple lists into hierarchical groups. Intra-group duplicates, however, are allocated randomly. Assume, for example, a list hierarchy of A/B, C/D/E:

- List A would be credited with all names appearing on lists A, C, D and E.
- Names common to Lists A, B, C, D and E, however, would be allocated equally among A and B.
- Likewise, names appearing only on Lists C, D and E would be allocated equally among the three.

This is a rational approach for established firms, where outside lists can be divided into two groups: proven lists and untested/unproven lists.

- The proven lists are combined into a primary hierarchy group, and the untested lists into a second, lower-priority group.
- For both, intra-group duplicates are allocated randomly.
- All overlap between the proven and the untested groups are allocated to the proven group. The untested group is credited only with incremental names not appearing in the proven list group. The underlying philosophy is that the proven lists would be mailed anyway. An unproven list is considered successful only to the extent that it makes a net contribution to the universe of economically mailable names.

This approach has the advantage of response reports that reflect the incremental value of new lists; a base from which rational rollout decisions can be made. Although it does not minimize rental charges, it is recommended for most established direct marketers.

- We believe that accurate rollout decisions far outweigh the associated incremental rental charges.
- These charges are small relative to total rental costs. Total rental costs, in turn, are a minor portion of total in-the-mail-costs.

Exhibit 1 illustrates the impact on response reports of random allocation within a hierarchy.

Separation of Duplicates

Some mailers separate duplicates entirely from the underlying lists, treating them as if they were a discrete rental list of multi-buyers.

- The disadvantage is that individual list performance reflects only single-buyer responses, and is therefore understated.
- The level of understatement is positively related to the percentage of duplicates. Understatement is most pronounced in lists with a high percentage of duplicates, which are likely to be your strongest performers.
- The percentage of duplicates almost always varies, which means the level of understatement will be different across lists.
- With different levels of understatement, comparison of results becomes distorted.

Exhibit 1 illustrates the impact on response reports of duplicate separation.

Exhibit 1

List Overlap

	Gross Names	List A	List B	List C	List D	List E	Unique Names
List A	10,000	N/A	1,000	1,000	1,000	1,000	6,000
List B	10,000	1,000	N/A	1,000	1,000	1,000	6,000
List C	10,000	1,000	1,000	N/A	1,000	1,000	6,000
List D	10,000	1,000	1,000	1,000	N/A	1,000	6,000
List E	10,000	1,000	1,000	1,000	1,000	N/A	6,000

Comparison of Various Duplicate Allocation Techniques on Reported Response Rates

	Random		Hierarchical ³		Random Within A Hierarchy ⁴		Separation of Duplicates		
	Unique Names	Dupe Allocation	Overall Resp Rate ²	Dupe Allocation	Overall Resp Rate ¹	Dupe Allocation	Overall Resp Rate ¹	Dupe Allocation	Overall Resp Rate ¹
List A	6,000	2,000	2.50	4,000	2.80	3,500	2.74	0	2.00
List B	6,000	2,000	2.50	3,000	2.67	3,500	2.74	0	2.00
List C	6,000	2,000	2.50	2,000	2.50	1,000	2.29	0	2.00
List D	6,000	2,000	2.50	1,000	2.29	1,000	2.29	0	2.00
List E	6,000	2,000	2.50	0	2.00	1,000	2.29	0	2.00
Dupes	N/A	N/A	2.50	N/A	N/A	N/A	N/A	10,000	4.00
Total	30,000	10,000	2.50	10,000	2.50	10,000	2.50	10,000	2.50

² List A to E unique name response rate = 2.00%. Duplicate response rate = 4.00%.

³ List hierarchy: A = highest; then B; then C; then D; E = lowest.

⁴ First list hierarchy = A & B. Second list hierarchy = C, D & E.

The List Matrix Report

Regardless of the allocation method employed, merge/purge results should be summarized in a List Matrix Report, as follows:

- Gross input quantities, by list.
- The absolute and percent overlap of each list with all other lists.
- Total duplicates “credited” to each list.
- The resulting net quantities per list.

Multiple Mailings of Duplicates

The following outlines how an established direct mail company could monitor the performance of duplicates that are mailed multiple times:

- Allocate according to the rules outlined in the random-allocation-within-a-hierarchy section.

- Append duplicate and mail drop suffixes to the main source code. Two-time duplicates in the first mail drop, for example, would receive a “-21,” three-time duplicates in the second drop a “-32,” etc.
- For each mail drop, track first by main source code and second by duplicate suffix.
- Response reports for each drop should be broken down in an identical manner, with main source code results separated from duplicates results in order to prevent double-counting.
- Ideally, duplicates response for each drop would be tracked by list (A-21, A-31...B-21, B-31...etc.), but in reality this may prove unwieldy. At a minimum, results should be aggregated by overall suffix (2X, 3X, etc.).

Ethically, a duplicate should be mailed multiple times only when each list owner has given prior consent, agrees to subsequent mail dates for multi-buyers, and has been compensated in full for the use of the name.

- Regardless, duplicates should never be mailed more times than they appear on outside lists (twice for 2X duplicates, three times for 3X, etc.) without the list owner receiving additional compensation.
- Net name rebates should not be expected when duplicates are mailed the same number of times they appear on outside lists.

Exhibit 2 illustrates how to track and report multiple mailings to duplicates.

**Exhibit 2:
Format Outline for Tracking Multiple Mailings to Duplicates**

	Primary Mailing			Re-mail #1			Re-mail #2 ⁷		
	Qty Mailed	Response Qty	Response Pct	Qty Mailed	Response Qty	Response Pct	Qty Mailed	Response Qty	Response Pct
List 1	10,000	500	5.00	1,144	36	3.15	485	12	2.47
List 2	10,000	300	3.00	1,144	37	3.23	486	13	2.67
List 3	10,000	100	1.00	1,145	37	3.23	486	13	2.67
	30,000	900	3.00	3,434	110	3.20	1,457	38	2.61
Duplicates⁵									
2x	2,000	80	4.00	1,920	53	2.76	0 ⁶	–	–
3x	1,000	50	5.00	950	33	3.47	917	22	2.40
4x	600	36	6.00	564	24	4.26	540	16	2.96
	3,600	166	4.60	3,434	110	3.20	1,457	38	2.61

⁵Duplicate numbers are reflected in the by-list mail report above.

⁶Not ethical to mail unless list owners receive additional payment, regardless of whether the mailing makes economic sense.

⁷The same format would be followed by Re-mail #3, where only the 4X duplicate group would be mailed (assuming this makes economic sense).

What to Look for In Evaluating Merge/Purge Systems

Over the past six articles, we have given you a great deal of specific information about the merge/purge process. Here, however, is the primary thing to remember:

Many direct marketing professionals have limited technical knowledge of the merge/purge process. This lack of knowledge is often compounded by an unwillingness or inability to communicate needs on anything but the vaguest of levels. This is known as the “black box” mentality. Or, some are quite involved in the initial selection of a vendor, and then move on to other matters, ignoring future merge/purge issues or needs.

These same professionals, however, devote much time and effort to the other aspects of a mailing. We hear much discussion of ZIP overlays, file segmentation, and list testing. We do not want to imply that these are not important, because they are. But, it is also important to realize that similar

attention to the merge/purge process can result in significant improvement in the quality of names mailed and the cost of doing so.

Jim Wheaton and Cynthia Baughan Wheaton are Principals at Wheaton Group, and can be reached at 919-969-8859 or jim.wheaton@wheatongroup.com. The firm specializes in direct marketing consulting and data mining, data quality assessment and assurance, and the delivery of cost-effective data warehouses and marts. Jim is also a Co-Founder of Data University www.datauniversity.org.