DATA FUSION IN THE BRITISH NATIONAL READERSHIP SURVEY - AN EXPERIMENT

Ron Carpenter - Mirror Group Newspapers
Steve Wilcox - RSMB Television Research Ltd

Summary

This paper reports on an experiment commissioned by the Press Research Council in Great Britain to examine the efficiency and quality of fusion as a method of transferring readership data collected from one sample onto the informants of another sample.

As subjects for the fusion, 176 magazines were chosen from the National Readership Survey. The principle was then to take one year's NRS sample, divide it into two halves, and transfer magazine readership data from one half to the other. Comparison of the fused results with the known 'real' readerships from the 'donor' and 'recipient' surveys would provide an exact measure of the accuracy of the fusion.
Amongst the fused magazines, the average issue readerships and internal duplications were very close to those on the original 'donor' survey. However, a final stage of ascription was also applied in order to bring average issue readerships exactly into line. Thus the prime objective to preserve the 'donor' survey trading currency was achieved.

The real test of the power of the fusion is to measure its ability to reproduce the duplication between individual variables on the 'recipient' survey (e.g. daily newspaper readership, product usage) and individual magazine readerships from the 'donor' survey. With a few exceptions, the fusion was also highly successful in this respect.

Part of the purpose of this experiment was to learn which kinds of magazine were most difficult to fuse efficiently. There were only a few cases arising, where the linking variables have not been as discriminatory as for the majority of publication groups. The solution, in principle, is to add further linking variables - either by using new combinations of the questions already asked on both surveys, or by designing additional hooks to go on both surveys in the future.

Comparing the two approaches to the fusion, we concluded that it is better to do a series of mini-fusions in which publications are fused block by block, one block at a time, than to fuse all publications in a single global fusion.

In conclusion, the representatives of the Press Research Council were re-assured that fusion could be used as a practical and effective alternative when a single source survey is not feasible or to combine two existing independent surveys.

1. Introduction

In 1991, the Press Research Council, which has representatives of all the major national newspapers and major consumer magazine publishers in Great Britain, wished to examine the efficiency and quality of fusion as a method of transferring readership data collected from one sample onto the informants of another sample.

At the time this experiment was being planned, the shape of the National Readership Survey from 1992 onwards was the subject of debate. There was a prospect that the NRS might split consumer magazines into two lists, with one list asked about on one half of the sample and the other list asked about on the other half of the sample. In this case it would be desirable for each magazine to have its readership fused onto the half of the NRS sample which had not covered the title. An alternative possibility was that some of the smaller magazines might be dropped from the NRS altogether. In this case, for a period after the removal, the past NRS readerships could be fused onto the latest period (adjusted if necessary by circulation or other relevant changes). For either of the circumstances, fusion of NRS data onto a different set of NRS informants was the appropriate test.

In the event, the arrangements for 1992 meant that the NRS has been able to accommodate a full range of magazines across the whole sample. Nevertheless this fusion experiment is still very relevant because the structure of future NRS contracts may require fusion and there a number of NRS-related circumstances in which fusion could be a valuable technique. The most obvious type of example is where the NRS is to receive data from a separate survey among the general population which asked in detail about a specific market area (such as informants' financial arrangements) and the readership of associated titles not covered by the NRS.

The broad objectives of the test were to learn more about fusion methodology in the context of the NRS and other readership surveys, to assess how accurate fusion can be, and to put publishers in a position to implement a 'live' fusion quickly and efficiently using tested techniques. A specific objective was the need to preserve on the recipient survey the **exact** readership levels found in the donor survey, because so much advertising revenue hinges on the relationship between the readership figures. Normally one could expect to get the fused readerships close to the donor levels but not exact. The solution was to carry out an ascription exercise after the fusion had been completed to bring the fused readerships exactly into line with the donor survey.

In order to progress the test of fusion, from its research committee the Press Research Council set up a working party, whose membership was as follows:

Guy Consterdine	Chairman
Ron Carpenter	Mirror Group Newspapers
Geoff Wicken	G & J of the UK

The working party appointed Ken Baker as fusion consultant and subsequently RSMB Television Research Ltd (represented on the working party by Steve Wilcox) to carry out the fusion. RSMB was chosen because of their experience in the field and proven track record in the fusion of the BARB TV audience measurement panel and the Target Group Index consumer survey.

2. The Fusion Experiment

As subjects for the fusion, 176 magazines were chosen from the 1990 NRS. The principle was then to take one year's NRS sample, divide it into two halves, and transfer data from one half to the other half. Comparison of the fused results with the known 'real' readerships from the 'donor' and 'recipient' surveys would provide an exact measure of the accuracy of the fusion.

January-June 1990 NRS was chosen as the 'donor survey', whose magazine readerships were transferred onto July-December 1990 NRS, the 'recipient survey'.

This is an ideal circumstance for fusion. The two surveys have a common universe, identical readership questions, the same classification and other 'link' variables, and large samples of equal size and identical structure. If two surveys that were very different from each other were fused one would not expect to produce as good a match from fusion alone as could be achieved in this experiment. However the use of ascription as a final stage would guarantee that real-life fusions achieved identical average issue readerships among the total universe.

In most fusions, account will have to be taken of differences in the sample profiles of the donor and recipient surveys, in terms of important linking variables such as age, social grade, sex and so on. For this test there was no need to do this, since we were using successive 6-month periods of the same continuous survey.

A basic principle of data fusion is that **all** the data collected from one informant in the donor sample is to be passed over to another informant in the recipient survey. Appropriate partners are chosen on the basis of their similarity in terms of a range of informant characteristics common to both surveys. For this fusion the common 'linking' variables included not only the obvious demographic ones but also some based on broad readership patterns:

Sex
Age
Social Class
Household Size
Head of Household
Housewife
Marital Status
Children 0-4
Children 5-15
Net readership of mid-market newspapers
Net readership of popular newspapers
Net readership of quality newspapers
Standard region
Terminal education age
Tenure type
WorkingStatus

Two fusion approaches were carried out. One was to fuse all of the 176 magazines simultaneously, in what we called the 'global fusion'. The other was to split the titles into blocks of magazines in the same field and carry out a separate fusion for each block.
For this purpose the 176 magazines were split into 33 blocks and a separate fusion was carried out for each of 12 of these blocks (in order to limit the size of the experiment but still covering 93 of the magazines):

Angling magazines
Motorcycle magazines
D.I.Y. magazines
Gardening magazines
Car magazines
Music magazines
Current affairs and finance
Fashionable women's magazines
Women's weeklies
Status magazines
General women's magazines
"Homes" magazines

Then, for example, the average issue readership and reading frequency data for Angling Magazines were removed from the recipient survey and replaced by a code indicating whether or not the informant had ever read at least one Angling Magazine. Thus a limited series or **block** of informants on the recipient survey were identified to be persons who at least sometimes read Angling Magazines. A similar process was conducted on the donor survey and thus **a future 'link' question or 'hook' between the parent survey (NRS) and independent surveys was simulated**. Then the fusion was confined to those recipients and donors within the Angling Block.

A further point to bear in mind is that there is more than one method of fusion. There is no single 'correct' method of fusion, nor are there rigid computer programs that can only conduct fusions in a fixed way. The exact means of implementing a fusion for a particular situation must be designed as a bespoke method for that situation. An outline of RSMB's fusion methodology is given in Appendix A.

3. Ascription

We saw it as important to preserve on the recipient survey the **exact** readership levels found in the donor survey, because so much advertising revenue hinges on the relationship between the readership figures. Normally one could expect the fusion to get the fused readerships close to the donor levels but not exact. For example the donor survey may show Magazine A with 2.0% penetration, Magazine B with 1.5% and Magazine C with 0.9%. If the result of fusion was that Magazine A was given 1.8% penetration, Magazine B 1.7% and Magazine C 0.7%, this would not be acceptable because the relationships between the magazines would have been changed so dramatically. We are well aware that all the above differences between the donor and fused figures for each title are likely to be within 95% confidence limits, but nevertheless the original figures on the donor survey remain the best estimates. Because the act of fusion introduces a further sampling error which compounds the sampling error already there in the donor survey, the probability is that the fused figures would be less close to reality. More importantly, the commercial situation means that such shifts in readership due to inexact fusing will not be acceptable.

The solution was to carry out an 'ascription' exercise after the fusion had been completed. Ascription adjusts the fused readerships **exactly** into line with the donor survey. This process is akin to the adjustments already made on NRS for circulation losses. It involves moving a few informants from one frequency claim to a neighbouring claim, and/or altering a handful of the readership probabilities associated with particular frequency claims.

Ascription is ruled out as the principal means of transferring data from the donor survey, because ascription cannot preserve the relationships between all the data. It would be a satisfactory method if all the relationships were governed by a single known characteristic, but where many characteristics are involved (as here) ascription would lose too much of the relationships.

4. Testing the Fusion

The objective of this experiment was to test whether or not fusion could be used to successfully marry the readership data from one survey to another. In this respect, it was not meant to be an evaluation of a particular fusion algorithm, in this case RSMB's. However, it is important to demonstrate that the fusion algorithm has performed well in matching the two surveys, given the information available. Then this can be eliminated as a potential source of error in the event that the fused database does not successfully simulate a

single source survey. So the fusion was tested first in terms of 'Fusion Diagnostics' and secondly in terms of 'Fusion Performance'.

We have already stressed the importance of preserving the 'Trading Currency' provided by a donor survey. This comprises the average issue readership of each magazine in the donor survey and the duplication between each pair of magazines, for a limited number of demographic groups (e.g. defined by region, age, class and sex). If all we had to do were to preserve these measurements in the fused database, then the fusion task would be relatively simple. In fact it would be based upon a stratified random sample of the donor survey. **Within each demographic group**, we would draw a random sample of informants from the donor survey and marry them at random to informants in the recipient survey within that demographic group. Any distortions in the fused magazine average issue readership and duplication figures (caused by not all donors being used exactly once) would be random and equivalent to a very small increase in sampling error. (This can only go 'wrong' if the recipient and donor surveys have significantly different sample structures. In this case, a decision would have to be made about the representativeness of the donor survey and what corrective weighting may be required).

Of course in practice we are trying to do much more than this. The difficult bit is to create a fused database which correctly measures the duplicate readership between each publication which is only carried in the recipient survey and each publication which is only carried in the donor survey. Pursuing the sampling analogue, we want to increase the stratification beyond a number of demographic groups in order to embrace all the relevant characteristics common to both surveys. This creates so many permutations of characteristics (far in excess of the survey sample sizes) that we are effectively stratifying at the individual informant level. For each informant in the recipient survey, the best we can do is to find the informant in the donor survey who matches most closely in terms of all the common characteristics. This gives the best chance of making an unbiased prediction of the magazine readership profile of each informant in the recipient survey. Hopefully this will lead to accurate predictions of magazine readership profiles for groups of informants with particular readership profiles from the recipient survey.

However, in creating this optimum demographic matching, there will be a great tendency to use some donors over and over again and not to use others at all. This is like drawing a smaller sample and then weighting it to match the characteristics of the recipient survey. This results in a loss in the effective sample size of the fused database and a greater potential for distortion in the 'passed over trading currency' - the magazine average issue readerships and duplications.

Therefore the fusion algorithm must optimise the matching process by trading off the conflicting objectives of:

(i) Maximising the degree of matching between recipients and donors = best chance of correctly predicting duplicate readership between a recipient survey publication and a donor survey publication.

(ii) Minimising the number of times each donor is used = best chance of preserving magazine average issue readerships and duplications from the donor survey.

This formulation demonstrates the direct link between the 'fusion diagnostics' and the 'fusion performance'. However, the fusion performance also depends upon the ability of the linking variables and linking function to explain all the systematic variability in the readership profiles in both surveys.

5. Fusion Diagnostics

The first half of the trade-off in the fusion optimisation is the degree of matching of recipients and donors.

In all the fusions, Sex and Broad Social Class (ABC1 vs C2DE) were defined to be critical cells and therefore recipients and donors match perfectly in all cases at this level. A typical example of the overall degree of matching is that produced for the fusion of 'Fashionable Women's Magazines' shown in Table 5.1.

Table 5.1 Fashionable Women's Magazines

Linking Variable	Number of Cells	% of Marriages Which Were:	
		Exact	Within ± 1 Cell
Sex	2	100	100
Class	6	72	99
Age	6	58	91
Household Size	5	41	83
Head of Household	2	93	100
Housewife	2	99	100
Married	2	98	100
Children 0-4	2	98	100
Children 5-15	2	94	100
Mid-Market Press	5	89	100
Popular Press	5	84	99
Quality Press	5	77	99
Standard Region	11	22	50
Terminal Education/Age	4	74	95
Household Tenure	2	97	100
Working Status	3	76	93

The relatively low success for Household Size reflects its level of importance as a determinant of readership of 'Fashionable Women's Magazines'. On the other hand, the apparent lack of success with standard region is actually due to the large number of cells. In reality the fusion has been successful in controlling the broad North, South, East and West balance.

On average, each married pair were matched perfectly on 13 of the 16 (13 of the 15 if we exclude standard region) criteria and matched within one cell on 15 of the 16 criteria. Table 5.2 lists these summaries for all fusions.

Table 5.2 Average Number of Characteristics Matched (Maximum 16)

Fusion	Exactly	Within ± 1 Cell
Angling Magazines	12	15
Motorcycle Magazines	12	15
D.I.Y. Magazines	12	15
Gardening Magazines	13	15
Car Magazines	13	15
Music Magazines	13	15
Current Affairs & Finance	12	15
Fashionable Women's Magazines	13	15
Women's Weeklies	13	15
Status Magazines	13	15
General Women's Magazines	13	15
'Homes' Magazines	12	15
Global Fusion	13	15

The second half of the trade-off is the number of times each donor was married. Table 5.3 shows the donor frequency distribution for each fusion compared with the distribution which would have been achieved had strict nearest neighbour matching been adhered to (i.e. always marrying the closest matches irrespective of how many times such a donor has already been used).

Table 5.3 Donor Frequency

Block		Recipient Sample	Donor Sample	Donor Frequency 0 %	1 %	2 %	3 %	4+ %
Angling	Nearest Neighbour	768	754(100%)	42	31	18	6	4
	Actual Matching	768	754(100%)	11	77	13	0	0
Motorcycle	Nearest Neighbour	887	800(100%)	38	33	17	8	5
	Actual Matching	887	800(100%)	7	76	17	0	0
D.I.Y.	Nearest Neighbour	742	807(100%)	45	31	15	6	3
	Actual Matching	742	807(100%)	18	72	10	0	0
Gardening	Nearest Neighbour	1453	1402(100%)	44	29	15	7	5
	Actual Matching	1453	1402(100%)	12	72	16	0	0
Car	Nearest Neighbour	2357	2371(100%)	45	30	15	7	4
	Actual Matching	2357	2371(100%)	14	73	13	0	0
Music	Nearest Neighbour	1952	2060(100%)	44	31	15	6	3
	Actual Matching	1952	2060(100%)	15	76	9	0	0
Current Affairs	Nearest Neighbour	1516	1024(100%)	12	29	18	10	10
	Actual Matching	1516	1024(100%)	0	54	43	3	0
Fashionable Women's	Nearest Neighbour	2497	2711(100%)	45	12	14	0	3
	Actual Matching	2497	2711(100%)	17	74	9	6	0
Women's Weeklies	Nearest Neighbour	6956	7014(100%)	46	28	15	6	5
	Actual Matching	6956	7014(100%)	17	67	15	1	0
Status	Nearest Neighbour	3838	4109(100%)	47	29	15	6	4
	Actual Matching	3838	4109(100%)	19	69	11	0	0
General Women's	Nearest Neighbour	4012	4236(100%)	46	30	16	6	4
	Actual Matching	4012	4236(100%)	12	69	12	0	0
Homes	Nearest Neighbour	2901	3137(100%)	47	29	14	6	4
	Actual Matching	2901	3137(100%)	20	68	11	1	0
Global Fusion	Nearest Neighbour	11217	11534(100%)	47	29	14	6	5
	Actual Matching	11217	11534(100%)	19	66	15	1	0

In general, the results show that about 70% of respondents were used once, 15% twice and 15% not at all. The incidence of donors used more than twice was negligible. This is much better than strict nearest neighbour matching which would have led to nearly 50% of donors not being used at all and consequently about 10% being used more than twice.

The distribution for the 'Current Affairs' block stands out as very different from the rest. This is because the penetration of this block grew by 50% from the first to second half of 1990. The resulting difference in sample sizes means that the **average** donor frequency must be 1½ compared to nearly 1 for all the other blocks. Of course this questions the validity of fusing out of date data in a growing sector. In this situation, careful consideration must be given to the application of the fused data.

The quality of the achieved donor frequency distribution means that the fusion must have successfully constrained the sampling error as far as the passed-on magazine average issue readerships and duplications are concerned. The price to pay is that marriages were not as close as they would have been under strict nearest neighbour conditions. However, as we have already seen, the average marriage is matched perfectly on 13 of the 16 criteria and within one cell on 15 of the 16 criteria.

One method of showing the loss of efficiency in matching is based upon the distance measurement (see Appendix A) which summarises the difference between a donor and recipient in terms of the common characteristics. Table 5.4 compares the sum of distances which would have been achieved under strict nearest neighbour conditions with the actual sum of distances achieved. Because distance measures have meaning only in comparative terms, the measures are presented in index form, the base for the index is the sum of distances which would have been observed by random allocation of donors to recipients.

Table 5.4 Donor Sum of Distances

	Sum of Distances (Index)		
	Nearest Neighbour	**Actual Matching**	**Random Matching**
Angling Magazines	12	15	100
Motorcycle Magazines	10	13	100
D.I.Y. Magazines	9	11	100
Gardening Magazines	7	10	100
Car Magazines	5	6	100
Music Magazines	9	12	100
Current Affairs & Finance	13	19	100
Fashionable Women's Magazines	8	10	100
Women's Weeklies	2	2	100
Status Magazines	1	2	100
General Women's Magazines	2	3	100
'Homes' Magazines	1	2	100
Global Fusion	6	8	100

The data show that distance measures for the actual fusion are only slightly higher than under strict nearest neighbour conditions. This means that in constraining the donor frequency distribution, we have not significantly affected the degree of matching between donors and recipients. Therefore the chances of correctly predicting the duplicate readership between a recipient survey publication and a donor survey publication have not been significantly reduced.

6. Fusion Performance

Evaluation of the performance of the fusion is in three parts:

- Preservation of the 'trading currency', i.e. how well are the donor survey results for the average issue readership of each magazine and their duplications replicated in the fusion data base?
- Effect of ascription, i.e. in using ascription to bring average issue readership exactly into line, how much are magazine duplications distorted.
- Accurate simulation of a single source survey, i.e. how well does the fused database measure the duplications of publications (and other classifications) on the recipient survey with magazines on the donor survey.

6.1 Preservation of the Trading Currency

In the global fusion, individuals' readership records for 176 magazines were passed over from the donor to recipient survey. In the block fusions, 93 magazines were split into 12 blocks and individuals' magazine readership records were passed over in 12 separate fusions.

Table 6.1a shows how well the fused database replicates the average issue readerships of a selection of 20 magazines, based upon the block fusions.

Table 6.1a Average Issue Readership (Adults)

	Donor Survey (i.e. the data the fused survey must match) %	Fused Survey before ascription %	Fused Survey after ascription %
Anglers Mail	0.7	0.7	0.7
Motor Cycle News	1.7	1.8	1.7
Do It Yourself	1.4	1.3	1.4
Garden News	0.9	0.9	0.9
Classic Cars	2.8	2.8	2.8
Motor Sport	1.7	1.8	1.7
Auto Express	1.9	1.9	1.9
New Musical Express	1.3	1.3	1.3
Smash Hits	4.5	4.2	4.5
The Economist	0.9	1.1	0.9
Investors Chronicle	0.3	0.4	0.3
Options	1.6	1.5	1.6
Vogue	3.6	3.4	3.6
Womans Own	9.9	9.9	9.9
Bella	10.2	10.5	10.2
My Weekly	3.8	4.0	3.8
Country Living	1.8	1.7	1.8
She	3.1	3.0	3.1
Living	1.6	1.6	1.6
Private Eye	1.7	1.7	1.7

These 20 magazines are representative in that the distribution of differences between the donor survey and the fused survey prior to ascription is the same as the distribution of differences across all magazines.

Among the 93 magazines in the block fusions, the fusion (before ascription) produced identical readerships to the donor survey for 38% of the magazines. 73% of magazines were within ± 0.1% and 90% within ± 0.2%. These results are excellent in themselves, but the final stage of ascription goes a step further in making the readership for every single title identical to its figure in the donor survey, as the final column in table 6.1a shows. Table 6.1b compares the full distribution of differences from the block fusions with that generated by the global fusion - the latter is slightly better although it must be noted that we are not strictly comparing like with like because the global fusion covers another 83 magazines.

Table 6.1b Average Issue Readership (Adults)
Distribution of Absolute Differences

Fused - Donor	Block Fusions (93 Magazines)	Global Fusions (176 Magazines)
+ 0.5 or greater	0%	0%
+0.4	0%	0%
+0.3	2%	1%
+0.2	9%	3%
+0.1	13%	13%
0.0	38%	46%
-0.1	23%	29%
-0.2	9%	5%
-0.3	7%	2%
-0.4	0%	0%
-0.5 or less	1%	1%

Table 6.1c gives a few examples of the way the fused database retains the readership duplication patterns found in the donor survey, based upon the block fusions.

Table 6.1c
The Preservation of Readership Duplications

Adults	Donor Survey %	Fused Survey* %
Readers of Anglers Mail who also read:		
- Angling Times	72.0	72.3
- Trout Fisherman	4.3	3.2
Readers of Management Today who also read:		
	20.2	19.5
- The Economist	3.2	2.4
- Time		
Readers of melody Maker who also read:		
	66.3	65.4
- NME	12.8	12.3
- Record Mirror		
Readers of Company who also read:		
- Cosmopolitan	59.2	60.4
- Tatler	6.3	6.5
Readers of Woman who also read:		
- Womans Own	76.0	76.9
- Peoples Friend	9.5	10.0
Readers of House & Garden who also read:		
	38.0	37.8
- Ideal Home	6.3	7.6
- House Beautiful		

*Before Ascription

For each magazine taken as a base, figures are shown for one magazine with relatively high duplication and another with relatively low duplication. This high low pattern is very closely reproduced on the fused database.

Table 6.1d compares the full distribution of differences with that generated by the global fusion.

6.1d Readership Duplications (Adults)
Distribution of Differences

Fused - Donor	Block Fusions	Global Fusion
+10.0 or greater	1%	0%
+5.0 to +9.9	3%	1%
+3.0 to +4.9	8%	3%
+2.0 to +2.9	7%	3%
+1.0 to +1.9	6%	14%
+0.1 to +0.9	19%	19%
0.0	4%	6%
-0.1 to -0.9	25%	22%
-1.0 to -1.9	17%	14%
-2.0 to -2.9	6%	11%
-3.0 to -3.9	3%	5%
-5.0 to -9.9	0%	2%
-10.0 or less	0%	0%

The differences in this table look bigger because the duplications are percentages of small percentages. In reality they too are very small and are much less than sampling error. It was felt to be unnecessary to use ascription for duplications.

For both aspects of the passed on data, the global fusion produced slightly better results than the individual block fusions. However, given the background sampling error involved, we do not believe the block fusions to be systematically worse.

6.2 The Effects of Ascription

The concern is that in using ascription to bring average issue readerships exactly into line, magazine duplications may be distorted. The first thing to note is that in practice very few records have to be amended to force consistency because the fused and donor average issue readerships only differ by sampling error. So there is not much potential for distortion in the duplications.

Taking Angling Magazines as an example there are the following discrepancies:

Table 6.2a Average Issue Readership (Adults)

	Donor	Fused	Donor - Fused
Anglers Mail	0.69%	0.68%	+0.01 = +1 informant
Angling Times	1.20%	1.24%	-0.04 = -6 informants
Trout Fisherman	0.66%	0.66%	0.00 = 0 informants
Trout & Salmon	0.35%	0.37%	-0.02 = -3 informants

Ascription involves adjusting the readership records for 10 informants, which has the following effect on the readership duplications:

Table 6.2b Duplications Before and After Ascription

	Donor Survey	Fused Survey		Donor - Fused	
		Before	After	Before	After
Anglers Mail	100.0	100.0	100.0		
& Angling Times	72.0	72.3	70.5	-0.3	+1.5
& Trout Fisherman	4.3	3.2	3.2	+1.1	+1.1
& Trout & Salmon	4.3	4.3	4.2	0.0	+0.1
Angling Times	100.0	100.0	100.0		
& Trout Fisherman	5.6	5.9	5.5	-0.3	+0.1
& Trout & Salmon	4.3	4.7	4.3	-0.4	0.0
Trout Fisherman	100.0	100.0	100.0		
& Trout & Salmon	26.1	28.9	26.7	-2.8	-0.6

This example shows that the quality of the duplication in the fused data is preserved after ascription, as we would expect.

6.3 Accurate Simulation of a Single Source Survey

This is the real test of the data fusion, namely the ability of the fusion process to recreate the true inter-relationship between any variable A in the donor survey and any variable B in the recipient survey. In most circumstances the true correlation between A and B would be unknown, but in this experimental situation the true correlation is known - it is already there on the recipient survey. Thus this part of the evaluation involves the cross analysis of fused magazine readership data with data such as daily newspaper readership, ownership of durables, etc. which was deemed to be the kind of data which would remain on the main survey should the fusion solution ever be adopted.

It is the natural tendency of data fusion to reduce discrimination by a process known as regression to the mean. In the experiment, it was possible to measure regression to the mean by cross analysing, for example, real daily newspaper readership from the recipient survey first with fused magazine readership and then with real magazine readership. The following hypothetical example illustrates the principle:

	% Who Read Magazine X	
	Real Data	**Fused Data**
All Adults	20%	20%
Daily Mirror Readers	30%	28%
Difference	+10	+8
	(100%)	(80%)

The fused data is understating the difference between the magazine readership of Daily Mirror readers and All Adults by 20% (i.e. 100% - 80%). In other words, there is 20% regression to the mean and the fusion is 80% efficient.

In the 12 block fusions, we are looking for regression to the mean in the readership of 93 magazines when cross-analysed by the readers of 27 newspapers/major magazines and 22 product usage type classifications; a total of about 4500 duplications.

In evaluating regression to the mean, we need to look at how many times it happens, how many times it is significant, how large it is on average and the extent in the worst cases.

If there were no systematic regression to the mean, then by chance we would expect it to occur in 50% of cases. The incidence we actually found was 57% amongst the readership classifications and 52% amongst product usage classifications. So as expected there is some systematic tendency for regression to the mean but certainly not all the time.

The next step is to determine whether or not each instance of regression to the mean is statistically significant. When making tests at the 5% level of significance, we would expect to exceed the tolerance in 5% of cases purely by chance. (This is the definition of the significance test). We actually found that the fused and actual duplications were significantly different in 10% of cases amongst the readership classifications and in 6% of cases amongst product usage classifications. So only a small proportion of instances of regression to the mean are statistically significant and almost all of these are amongst the readership classifications. However, as we all know, statistical significance is the last resort of the scoundrel! It is still important to measure the amount of regression to the mean.

The average amount of regression to the mean was only 26% amongst the readership classifications and just 1% amongst the product usage classifications. So on average, the fusion is highly efficient at measuring the inter-relationships between variables on the recipient survey and magazine readerships on the donor survey. But of course it is important to search for the worst cases in order to fully understand how well the fusion has worked. Table 6.3a shows the average amount of regression to the mean for each block of magazines on the donor survey.

Table 6.3a Regression of the Mean Recipient Survey Classifications

Magazine Block	Readership	Product Usage
Angling	23%	-16%
Motorcycles	22%	-13%
D.I.Y.	7%	-17%
Gardening	28%	6%
Cars	31%	-7%
Music	22%	-4%
Current Affairs/Finance	22%	10%
Fashionable Women's	18%	4%
Women's Weeklies	43%	20%
Status	18%	4%
More Women's	29%	17%
'Homes'	22%	-9%

[NB: A negative amount of regression to the mean indicates that the fused duplication is further from the mean than the actual. This is nothing more than a realisation of sampling error.]

The worst case is for Women's Weekly Magazines where on average there is 43% regression to the mean amongst the readership classifications and 20% amongst the product usage classifications. In the case of the readership classifications, this is largely explained by the fact that 4 of the 10 magazines in the block are also in the list of readership classifications from the recipient survey. In other words we are partly testing communality, i.e. the percentage of **individuals** for whom the fused and actual readerships of 4 magazines are the same. Fusion was never intended to perform well at this level. Whilst fusion can accurately predict the

magazine readership of the average Daily Mirror reader, it cannot accurately predict the magazine readership of each individual informant. If we remove these 4 magazines from the recipient survey, then regression to the mean amongst readership classifications would reduce from 43% to 27%.

But this example is useful in that it highlights the more general problem of splitting a group of similar magazines between the recipient and donor surveys. It would be much better to transfer all of them from the donor survey, thus ensuring that internal duplications are preserved.

Table 6.3b shows the average amount of regression to the mean of all 93 magazines within each readership/product usage classification on the recipient survey. Regression to the mean was in excess of 40% for readers of the Financial Times, Radio Times and TV Times, where we have learnt that additional hooks would be desirable. In the case of the Financial Times, the majority of the problem is caused by the failure to reproduce the true duplications with individual titles in the Current Affairs and Financial Magazines block.

Regression to the mean was also in excess of 40% for readers of the Daily Record. The Daily Record is only available in Scotland. The problem is that the structure of the regional hook allowed individuals in Scotland to fuse with individuals in the North of England. This has led to significant regression to the mean for those magazines (e.g. the Angling Block) which also have a strong regional bias. Therefore, whilst the regional hook has been adequate for the majority of publications and other classifications, we would recommend that in any future fusions standard region should constitute a critical cell. Then individuals in Scotland (for example) would always be married to individuals in Scotland.

Overall, regression to the mean was very low, particularly amongst daily newspapers and product usage groups.

Table 6.3b Regression to the Mean

Recipient Survey Classification Daily Newspapers	All Magazines
Daily Express	16%
Daily Mail	22%
Daily Mirror	10%
Daily Record	42%
Daily Star	0%
Daily Telegraph	19%
Financial Times	42%
Guardian	15%
Independent	15%
The Sun	6%
The Times	18%
Today	18%
Average	**18%**
Other Newspapers	
Racing Post	-25%
Sporting Life	-19%
Evening Standard	34%
Wolverhampton Express & Star	-19%
Newcastle Evening Chronicle	-6%
Average	**8%**
Major Magazines	
People Magazine	6%
Observer Magazine	31%
Radio Times	42%
TV Times	41%
Woman's Own	32%
Woman's Weekly	30%
Woman	29%
Woman's Realm	36%
Readers Digest	28%
Family Circle	33%
Woman & Home	28%
Average	**32%**

Recipient Survey Classification Product Usage	All Magazines
Income <£10k	-1%
Cinema Light	-34%
Cinema Medium	11%
Cinema Heavy	21%
Car 1	-7%
Car 2+	10%
Freezer	-10%
Lawn Mower	-2%
Microwave	19%
Home Computer	-2%
Holiday 1 a year	-16%
Holiday 2 a year	1%
Holiday 3+ a year	4%
Stocks & Shares	9%
Advertising Viewing	-8%
Wine 1+ a week	7%
Holidays GB	0%
Package Holiday	-5%
Business Flight	6%
Pregnant	13%
Degree Qualification	-2%
Credit Card User	1%
Average	**1%**

Finally, table 6.3c shows the amounts of regression to the mean for each magazine block based upon the **global** fusion. The block fusion regression to the mean figures are shown in brackets for comparison.

Table 6.3c Regression to the Mean - Global Fusion
Recipient Survey Classifications:

Magazine Block	Readership		Product Usage	
Angling	24%	(23%)	15%	(6%)
Motorcycles	-5%	(22%)	-4%	(-13%)
D.I.Y.	30%	(7%)	-12%	(-17%)
Gardening	38%	(28%)	19%	(6%)
Cars	25%	(31%)	21%	(-7%)
Music	29%	(22%)	17%	(-4%)
Current Affairs/Finance	28%	(22%)	24%	(10%)
Fashionable Women's	31%	(18%)	24%	(4%)
Women's Weeklies	48%	(43%)	28%	(20%)
Status	27%	(18%)	17%	(4%)
More Women's	55%	(29%)	34%	(17%)
'Homes'	48%	(22%)	18%	(-9)
All	**34%**	**(26%)**	**20%**	**(1%)**

This demonstrates that the individual block fusions are significantly more efficient than the global fusion in terms of recreating readership duplications and the magazine readership profiles of product usage type classifications.

7. Conclusions

The fusion algorithm performed successfully, with a very high efficiency of matching appropriate informants from the two surveys. In marrying donor informants to recipient informants partners matched each other on almost all the linking variables. At the same time, the great majority of informants from the donor survey were married to an informant from the recipient survey only once; less than 1% were married more than twice. This meant that the fusion algorithm itself could be eliminated as a potential source of failure.

The fused database contains readership data for all the 176 fused magazines, in addition to the newspapers and major magazines it already contained and which were not fused. Amongst the fused magazines, the average issue readerships and internal duplications were very close to those on the original 'donor' survey. However, a final stage of ascription was felt necessary in order to bring average issue readerships **exactly** into line. Ascription did not dilute other relationships (such as readership duplications) because only very small numbers of individuals needed to have their data adjusted. Thus the prime objective to preserve the 'donor' survey trading currency was achieved.

The real test of the power of the fusion is to measure its ability to reproduce the duplication between individual variables on the 'recipient' survey (e.g. daily newspaper readership, product usage) and individual magazine readerships from the 'donor' survey. With a few exceptions, the fusion was also highly successful in this respect. Overall, the fused database suffered from about 15% regression to the mean which means that 85% of the discriminatory power of a true single source survey was generated by the fused database.

Part of the purpose of the experiment was to learn which kinds of magazine were most difficult to fuse efficiently. These arise where the linking variables have not been as discriminatory as for the majority of publication groups. The solution, in principle, is to add further linking variables - either by using new combinations of the questions already asked on both surveys, or by designing additional hooks to go on both the donor and recipient surveys in the future. One example concerns Radio Times and TV Times where the demographics and newspaper reading characteristics in the list of common variables do not by themselves identify readers of each of these titles very efficiently. An additional characteristic is needed to provide the missing link, e.g. weight of TV viewing or claimed channel share.

A similar problem occurs when publications within a particular type of magazine are split between two surveys. Whilst the duplications with publications outside the magazine type are reproduced well, the internal duplications are not. If this has to be done then it is recommended that the subset of titles from the magazine type carried in the recipient survey are also carried in the donor survey in order to provide an efficient link.

Comparing the two approaches to the fusion, we concluded that it is better to do a series of mini-fusions in which publications are fused block by block , one block at a time, than to fuse all publications in a single global fusion.

This experiment emphasises the value of building into surveys which are likely to be fused, questions or hooks which may be valuable as linking variables - not only the obvious demographics, but also questions tailor made for the purpose of fusion. An example is questions asking if informants ever read publications of certain specific categories.

And finally, the experiment shows that the practical steps taken in setting up the framework for the fusion are as important as the sophistication of the fusion algorithm itself.

Appendix A The Fusion Algorithm

TECHNICAL CONSIDERATIONS

(1) The Sample Structures - A Comparison

The ideal situation in which a fusion could take place is one in which the universes of the recipient and donor surveys were identical, questions relating to common variables were identical, and sample structures and methodologies for both donor and recipient survey were very similar. In our experiment, fusing two halves of the NRS obviously fulfilled these conditions and it is to be hoped that should fusion be undertaken on future readership surveys, such considerations would be largely fulfilled.

(2) The Common Variables

The common variables used on this survey were as follows:-

- Sex
- Age
- Social Class
- Household Size
- Head of Household
- Housewife
- Marital Status
- Child 0-4
- Child 5-15
- Readership of mid-market press
- Readership of popular press
- Readership of quality press
- Standard region
- Terminal educational age
- Tenure type
- Working status

In practice the variables were divided into two types:-

(a) Critical variables - which must match exactly.

(b) Matching variables - which influence the algorithm, but may not match exactly.

The critical variables chosen were sex and class. Critical cells were:-

- Men ABC1
- Men C2DE
- Women ABC1
- Women C2DE

Thus Male ABC1's could only marry Male ABC1's..... and so on. In effect each fusion was conducted four times , within the four critical cells. However class was also used at the more detailed level as a matching variable within each of these fusions.

(3) The Distance Measure

A good fusion algorithm must contain a distance measure which is:-

(a) Scale free

(b) Takes account of the degree of correlation between the common variables.

The distance measure chosen in the RSMB algorithm is Mahanalobis'distance. The distance between two individuals in terms of the common variables is given as:-

$$D^2 = W^T . (X_1 - X_2)^T V^{-1} (X_1 - X_2) . W$$

Where
- W is a vector of importance weights
- X1 is a vector of common variables for individual 1
- X2 is a vector of common variables for individual 2
- V is the variance /covariance matrix for the common variables as estimated in the recipient sample.

This highly complex matrix algebra equation can be literally translated as follows. The squared distance between individuals 1 and 2 is equal to the sums of squares of their distances on the common variables, adjusted for the degree of distances on the common variables, adjusted for the degree of correlation between the common variables. In addition, the individual components of the equation (the common variables) are weighted separately according to their degree of importance in discriminating between products and time segments.

Mahanalobis'distance is rendered scale free by division by the variance of each common variable. Thus common variables with a high variance due to scaling e.g. age, are equated to common variables with a low variance e.g. presence of absence of Child 0-4. Thus the distance measure will not favour age at the expense of presence of a 0-4 year old purely as a result of the scale used. (In practice age was often determined to be a better discriminator than presence of a 0-4 year old for many types of magazines, and was thus given a higher importance weight. However, here the matching procedure was influenced by the analyst, and not the scaling effect.)

A good distance measure must also avoid multi-counting of inter-correlated variables. For example, there is a high correlation between presence of children and household size. Thus there is an extreme likelihood that individuals who differ markedly on household size will also differ markedly on presence of children. However, doubling the distance because of high correlation would place undue emphasis on differences due to household structure; as against difference in uncorrelated variables, which would only be counted once.

Mahanalobis'distance accounts for this by dividing each of the components of distance (common variables) by its covariance (a measure of correlation) with all other variables.

(4) **The Importance Weights**

As outlined in previous sections, the algorithm employed differential weights of importance for the common variables. Distances between variables deemed to be important were highlighted by these weights relative to those of lesser importance. The algorithm thus tried to influence those variables for which the best possible matching was desired.

The method of determination of the importance weights was analysis of variance (ANOVA). This technique highlights the discriminatory power of each common variable on the variables to be donated, and the variables already on the recipient survey, in terms of a measure called the mean square. The mean squares of each common variable were then summed and expressed in terms of an average mean square - an indicator of the average discriminatory power.

This average mean square was calculated separately for data derived from the donor file, and data derived from the recipient file. A common variable which is highly discriminatory on the variables on the donor file, but not at all discriminating on the recipient file is in reality an unimportant variable in the fusion. If, for example, all newspapers have the same class profile, then each one would be expected to have approximately the same penetration of diamond ownership. However, diamond ownership is heavily skewed in terms of class. Thus in a fusion between a newspaper readership survey and a survey assessing ownership and likelihood of buying diamond jewellery, class is a highly discriminatory variable on one survey, and not at all discriminatory on the other survey. Simple random allocation on class will recreate the classless profile of newspaper readership and the lack of discrimination on diamond ownership. Hence, because random allocation on class variables will recreate reality, class is a totally unimportant variable in the fusion.

Thus the importance weights should reflect the degree of discrimination of each common variable amongst variables in both donor and recipient survey. In practice the mean squares produced by ANOVA from data on both surveys are multiplied together, and the square root of the product renders the importance weight.

(5) **The Marriage Algorithm**

Once the distance between any two respondents in the donor and recipient surveys have been obtained, the algorithm uses the very simple nearest neighbour or least distance decision when marriages are made. As with most fusion algorithms, the intention is to limit as far as possible the number of times each donor is used, as multiple use of respondents will reduce the effective sample size for the sample drawn from the donor file and given to recipients. Thus, once a donor has been married a penalty weight is placed on that donor as follows (the symbol X used below represents the distance between a given donor and potential recipients).

Penalty After Donor Used Once	**- (X + 0.1) x 2**
Penalty After Donor Used Twice	**- (X + 0.1) x 4**
Penalty After Donor Used Three Times	**- (X + 0.1) x 12**