

Maria Silvia Farina
D O X A
Milan - Italy

"NEW EXPERIENCES ON MERGING IN ITALY"

1) OBJECTIVES OF THE PRESENTATION

Data fusion or merging is, at present, a well known and frequently applied technique.

Sometimes it works on the individual level, sometimes on the group level, with different neighbour definitions and different distancing concepts.

The topic of this presentation is not so much to illustrate how the Italian media-press fusion works in terms of a detailed description of the fusion process; rather we would like to emphasize the use of special variables involved in defining the similarity between individual subjects.

In fact, the Italian media-press fusion process principally completes information present in a reduced form for the media-press audience and marginally resolves problems of missing data.

It is really important to understand the general situation in which we started, some years ago, with experimental fusion on media-press data.

Also at present, the fusion process works on different media-press data, with the intent of furnishing a single source for data of magazines and newspapers.

2) HISTORY OF ITALIAN MEDIA PRESS SURVEYS

Since 1985, in Italy we have had regularly two independent surveys evaluating media-press audiences on a national basis, through personal interviews organized in two waves per year :

- ISPIPRESS (ISPI in the past) for weekly and monthly magazines together on the same questionnaire
- ISEGIPRESS (ISEGI in the past) for daily newspapers.

NEW EXPERIENCES ON MERGING IN ITALY

In 1986 we started fusing data on magazine audience, coming from 2 different sources :

- the traditional ISPI survey (15,500 cases)
- the new media-press survey NIP(12,000 cases),collecting information on a subset of magazines surveyed with ISPI.

The main objective of this first fusion experiment was to enrich the basis for readership evaluation,extending the number of magazines available for media planning as well.

We had also an indirect advantage, because the validity of the results convinced the publishers of the possibility of gathering different reading and readership information from different samples of interviewees (not making them answer a long list of questions, taking-up a lot of time) without losing the possibility to use data together in planning processes.

The ISPI Technical Committee decided to start in 1987 with a new questionnaire to collect complete information on monthly magazines for a half sample and on weekly magazines for the other half-sample.

We moved from the first fusion experiment to fuse data on magazine audience using a methodology that still works, having been modified through the years with meaningful improvements.

In 1988, the first experiment took place for the fusion between data on national/local daily newspapers and data on magazines.

We attributed magazine reading information to 26,000 people from the ISEGIPRESS sample.

In the summer of 1990, we successfully concluded the realization of a complete common source for magazine and newspaper audiences, where the joined samples of media-press surveys are able to predict magazine and newspaper readership.

We enriched the magazine reading information of the 26,000 people from ISEGIPRESS sample and the newspaper reading information of the 25,000 people from the ISPIPRESS sample.

During the spring of 1990 a very important event took place: the creation of AUDIPRESS, a new partnership among press publishers and advertising users.

In accordance with the idea of a common source of information, the

NEW EXPERIENCES ON MERGING IN ITALY

combined ISPIPRESS-ISEGIPRESS Technical Committees decided to promote an experimental survey able to evaluate, at the same time, media press audiences and information about products and brands.

We are now working toward the completion of this ambitious project, whose complexity does not consist of the ability to complete weak information collected interviewing a large proportion of people, rather than to make statistically equivalent audience estimates evaluated on different samples, asked in a different way.

The experimental survey is being done, at the moment, on a small-sized sample (about 2400 people) representing two Italian regions, using a questionnaire quite different from the traditional one.

In fact, we have:

- 1) traditional part of questionnaire
 - weekly magazines
 - monthly magazines
 - daily newspapers
- 2) new part of questionnaire (self-administrated)
 - products and brands.

Other important differences between traditional and test surveys concern principally the internal disposal of questions inside the questionnaires and the interview flow.

As the traditional media-press survey and the experimental one run at the same time (they have a common two-week field work period) we will be able to evaluate the presence or absence of significant differences between audiences produced independently by the two surveys.

We think we will be able to extend in one year the actual experience to a complete survey of more than 50,000 people (25,000 for each wave, with two waves per year).

NEW EXPERIENCES ON MERGING IN ITALY

3) TRADITIONAL MEDIA PRESS SURVEYS

CHARACTERISTICS OF TRADITIONAL ITALIAN MEDIA PRESS SURVEYS

Presently we have two media-press surveys running 2 times per year, with the following characteristics :

- ISPIPRESS (about 12,500 cases per wave) - (Figure 1) -
with complete or "strong" information (7 questions) on :
 - about 40 weekly magazines
 - about 80 monthly magazines
 - 2 newspaper supplements

- with incomplete or "weak" information (2 questions) on :
 - about 50 national/local daily newspapers

- ISEGIPRESS (about 13,000 cases per wave) - (Figure 2) -
with complete or "strong" information (7 questions) on :
 - about 50 national/local daily newspapers
 - 2 newspaper supplements

- with incomplete or "weak" information (2 questions) on :
 - about 40 weekly magazines
 - about 80 monthly magazines.

The surveys are independent, but they have a common sample random design inside geographic strata.

They cover, also with over-sampling if it is necessary, small geographic strata corresponding to the areas of local newspaper sale.

The audience evaluation comes from the average of the most recent two waves of the surveys.

The two media-press surveys have different questionnaires but they have common questions for the socio-demographic part.

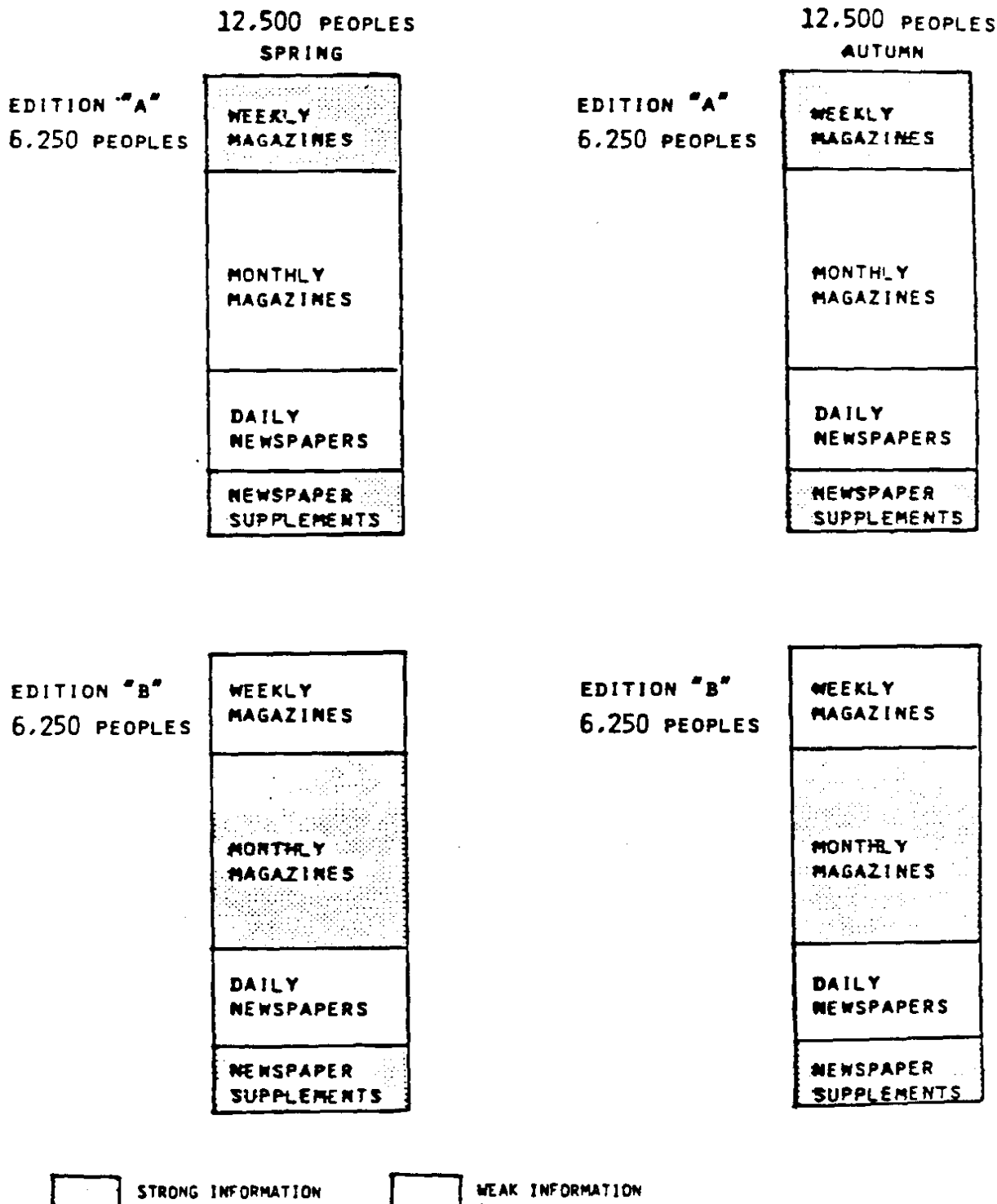
Inside each survey we have two editions of the same questionnaire:

- for ISPIPRESS
 - edition A with complete information about weekly magazines and reduced information about monthly magazines
 - edition B does the reverse

NEW EXPERIENCES ON MERGING IN ITALY

FIGURE 1

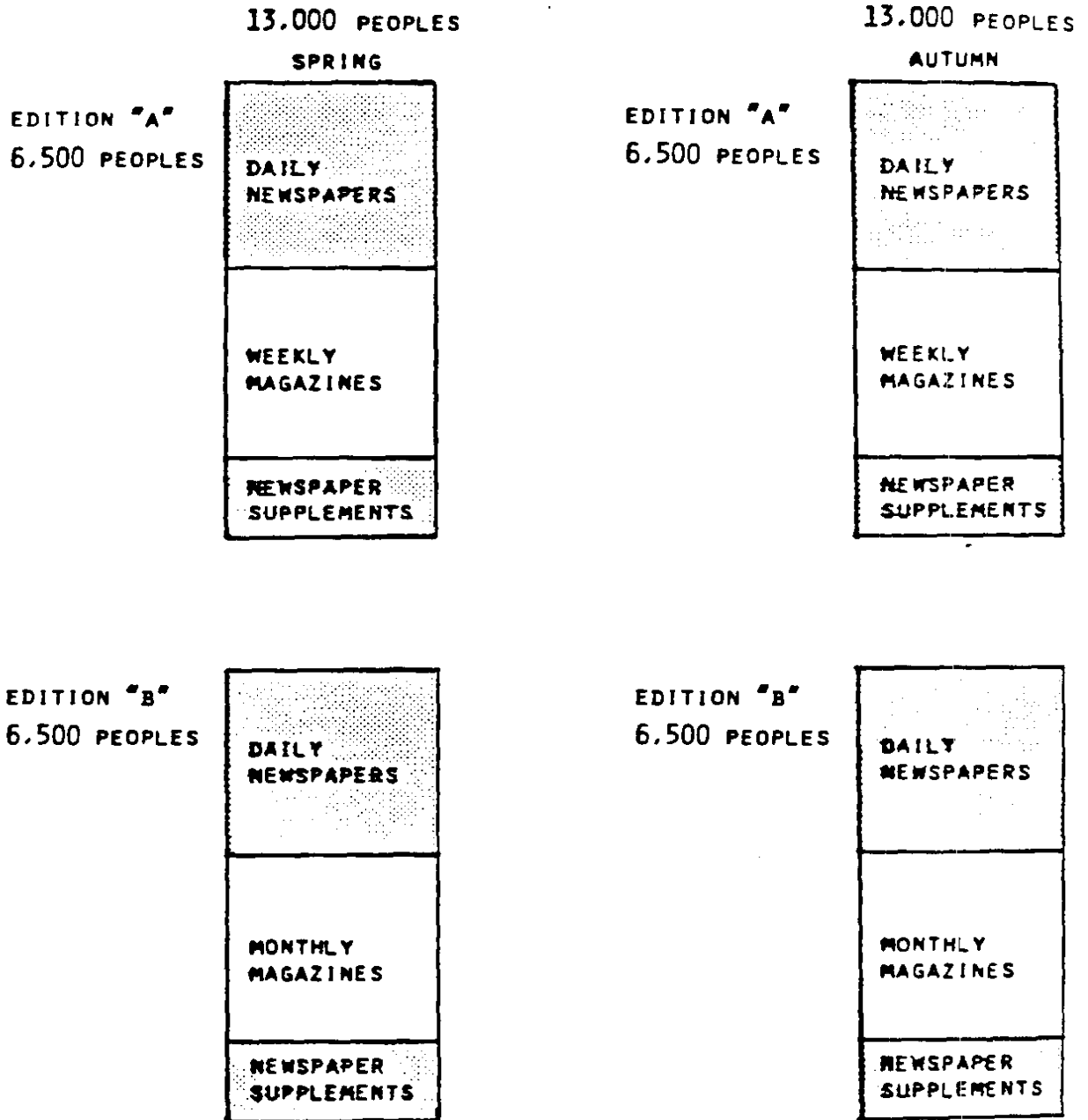
ISPIPRESS (MAGAZINES)





NEW EXPERIENCES ON MERGING IN ITALY

FIGURE 2

ISEGIPRESS (DAILY NEWSPAPERS)

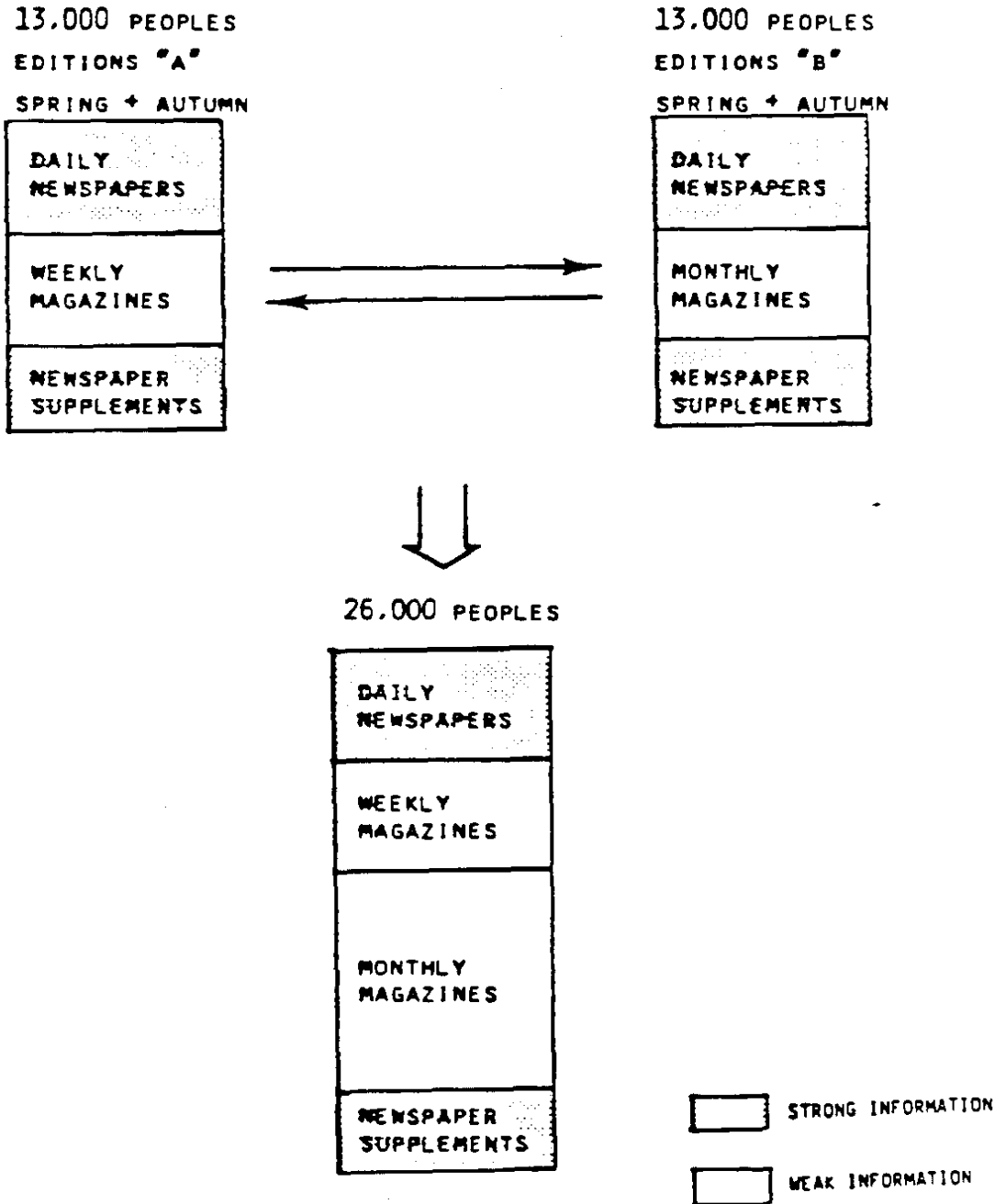


 STRONG INFORMATION
 WEAK INFORMATION

NEW EXPERIENCES ON MERGING IN ITALY

FIGURE 2 BIS

ISEGIPRESS (DAILY NEWSPAPERS)



NEW EXPERIENCES ON MERGING IN ITALY

both editions A and B have reduced information about daily newspapers and complete information about newspaper supplements;

- for ISEGIPRESS

edition A with complete information about daily newspapers and reduced information about weekly magazines

edition B for complete information about daily newspapers and reduced information about monthly magazines

both editions A and B have complete information about newspaper supplements - (Figure 2bis) -

The evaluation of the media-press audiences comes from the samples weighted independently, but using the same weighting cell definition.

MEDIA PRESS DATA FUSIONS

Two main fusions processes take place on media-press surveys :

- F1 fusion between data on magazines with different publication periods - (Figure 3) -
- F2 fusion between data on magazines and data on newspapers - (Figure 4) -

The fusion F1 has been active since 1987, while F2 started in 1988.

The objective of the fusion process F1 is to evaluate the duplicate audiences between weekly magazines and monthly magazines, allowing joint planning processes.

The objective of the fusion process F2 is to evaluate the duplicate audiences between weekly or monthly magazines and newspapers, allowing joint mixed planning processes.

The final result of media-press fusions is one file containing about 51,000 people which have complete information on :

- socio-economic variables (directly available from the 51,000 interviews)
- press audience variables (available from weekly magazines interviews, directly for 12,500 people of ISPIPRESS edition A sample, from monthly magazines interviews, directly for 12,500 people of ISPIPRESS edition B sample, from daily newspaper interviews, directly for 26,000 people of ISEGIPRESS sample edition A plus edition B).

NEW EXPERIENCES ON MERGING IN ITALY

FIGURE 3

ISPIPRESS (MAGAZINES)

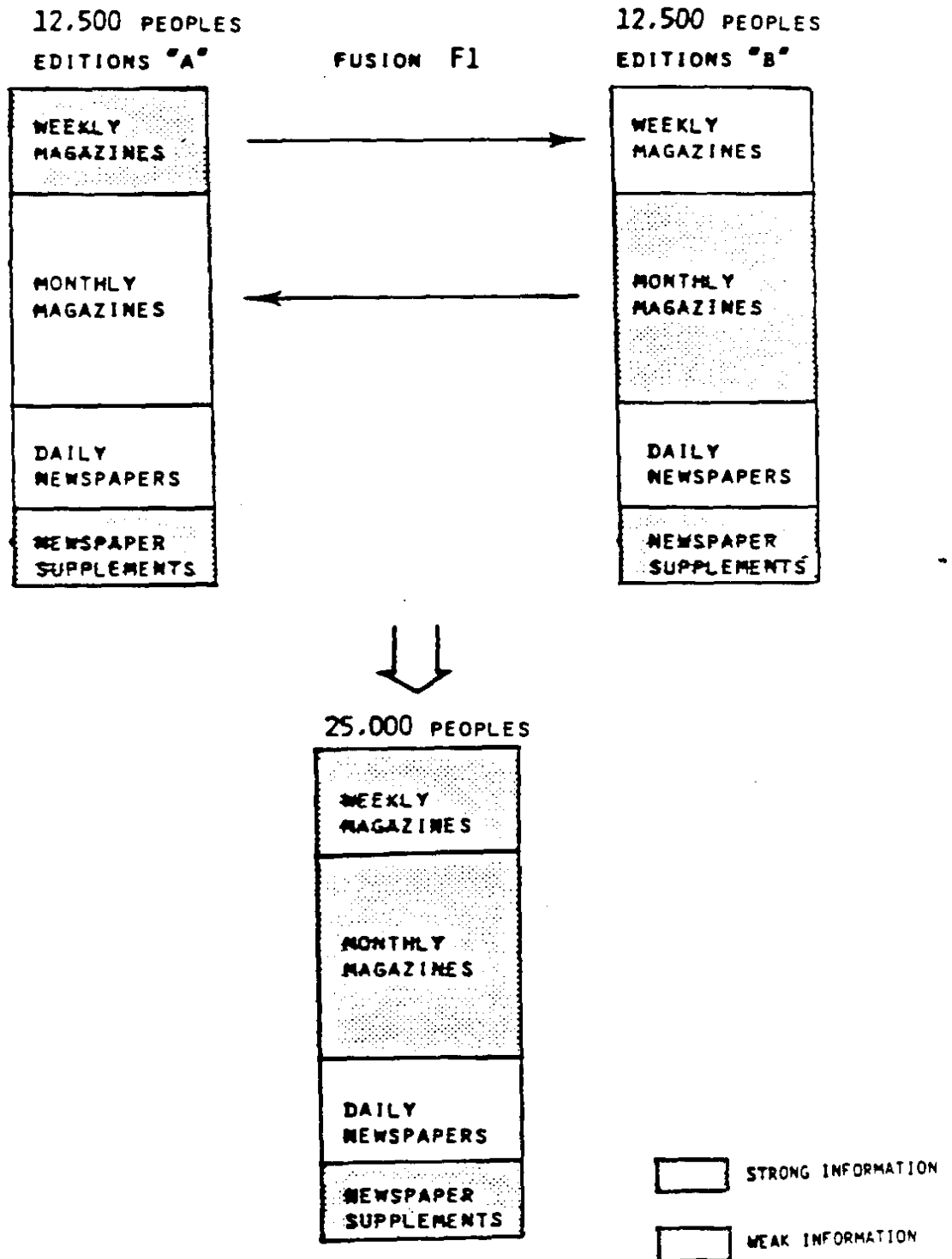
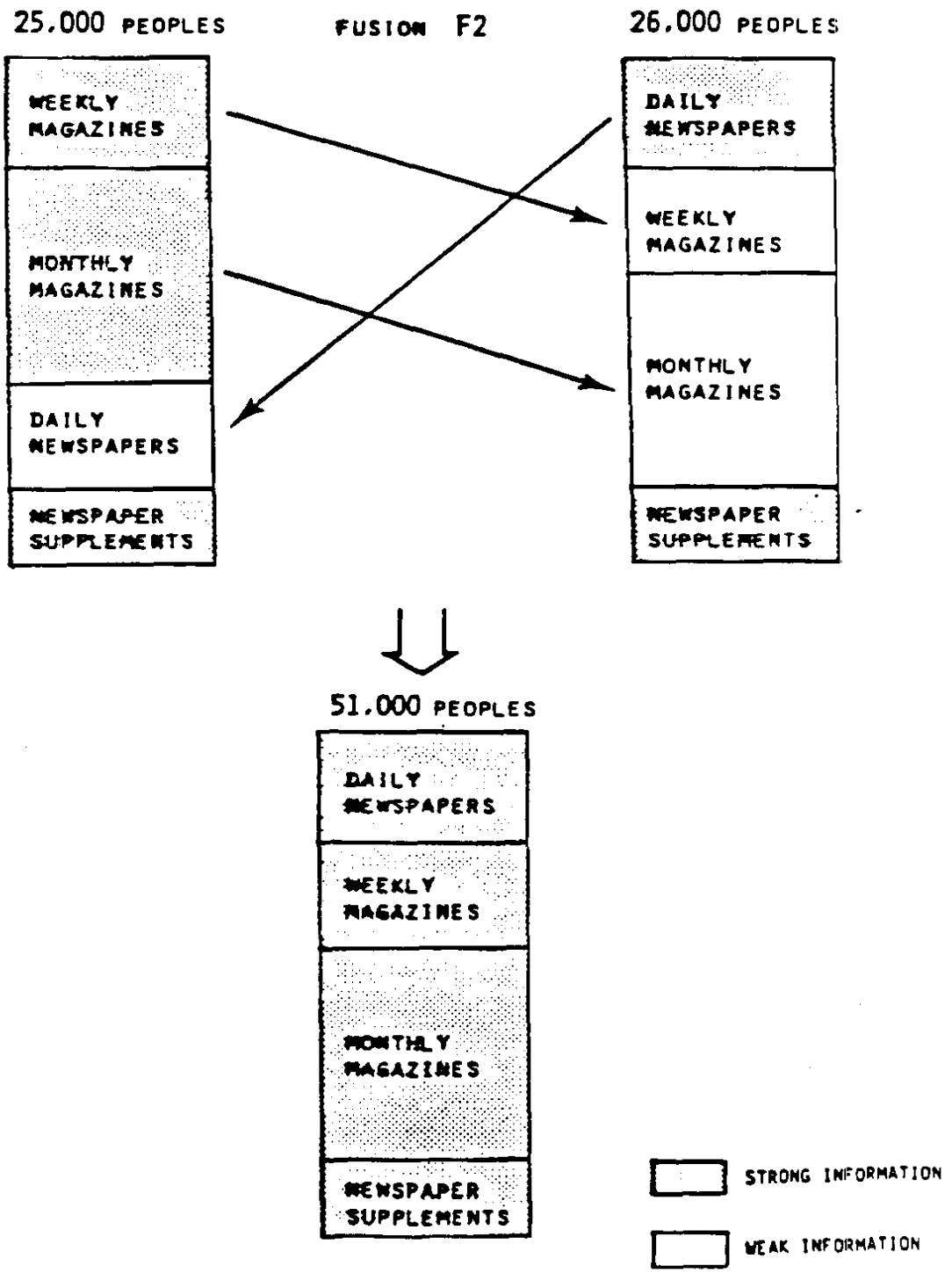


FIGURE 4

ISPIPRESS - ISEGI PRESS FUSION PROCESS



NEW EXPERIENCES ON MERGING IN ITALY

Each fusion process works on an individual basis, moving information from donors to recipients.

In fact, members of a survey or edition are matched with members of another survey or edition whose characteristics (demography, behaviour, ecc.) are closest.

The result of the individual matching process is the attribution of magazine/ newspaper reading from the donors to the recipients.

Two statements represent the basis for both processes:

- 1) the audience of each publication is derived from the sample collecting information in the more complete form;
- 2) the main task in the fusion process is the search for a recipient who must be the most similar possible to the donor, principally in the reading profile.

In practice the matching process takes place not only on general characteristics, but also works strongly at the reading level, using the weak reading information of the recipient sample like a magnet for moving the complete reading information from the donor sample to the recipient sample.

The choice of the variables active in the fusion process comes from a previous study searching for the socio-demographic characteristics that most discriminate people with different media-press audience behaviour (quantity, quality and groups of publications read).

The cells within which the fusion processes take place are defined by the following variables :

- first level cells
 - 1) small geographic areas (about 60)
 - 2) sex
 - 3) household responsibility
- second level cells
 - 4) age (8 classes)
 - 5) education level (5 classes)
 - 6) socio-economic status(5 classes).

Variable number one is involved in each fusion process because of the need to predict duplications between local newspapers and magazines at local geographic level.

NEW EXPERIENCES ON MERGING IN ITALY

The way we use the different kind of cells is regulated by a lot of norms, among which the most immediate and simple are the following:

- the first level cells represent strong constraints without possibility to go out of them

- the second level cells represent soft constraints where contiguous cells exchanges are allowed.

The reference system in which the similarity between people is measured is the same for both fusion processes and it uses the media-press audience variables common to both surveys in the weak form.

The similarity is measured by the "Euclidean distance".

The major part (65 %) of individual matches attribute reading information already present in the weak form in the recipient sample.

The remaining 35 % of matches take place only through similarity concepts.

EVALUATION OF THE FUSION RESULTS

At the end of the fusion processes, we plan several general checks.

The final figures for publication audiences are compared with initial figures coming from the sample collecting complete information on reading.

Chi-square analyses are provided on donor, donor joined to recipient samples for socio-demographic characteristics crossed with media-press audiences measured on single magazine/newspaper or group of publications.

Other checks are made on data before and after fusions, principally on overlapping between publications for which we have complete initial information.

All the previous checks took place at the end of the most recent fusion experiment, except the part concerning newspapers for which we had the complete original information.

For magazine audiences (reading in the last period), a good consi-

NEW EXPERIENCES ON MERGING IN ITALY

stency has been obtained between observed and after fusion results (see table 1).

Comparisons of magazine audiences cross-tabulated with socio-demographic variables, before and after fusion, don't show significant differences through CHI-SQUARE (see tables 2 and 3).

Finally, we compared publication audiences cross-tabulated with publication group audiences, both before and after fusion, through Chi-square.

Also in this case, we obtained a good consistency (see tables 4 and 5).

CONCLUSIONS AND PERSPECTIVES

During the execution of the last media-press fusion processes, we improved the following aspects of phases F1 and F2 :

- over-sampling treatment, when they are not equally present inside donor and recipient samples
- donors re-usage
- first level cell definition for both phases
- control of publications with low-level audiences.

We can conclude that, at the present, the media-press fusion processes are well-established procedures, able to supply media-press planners with consistent results.

NEW EXPERIENCES ON MERGING IN ITALY

TABLE 1

COMPARISONS

between

ISPIPRESS

ISEGIPRESS

after fusion

- WEEKLY MAGAZINES
concerning the 41 publications studied, the discrepancies
between the audiences measured by the two files are :
 - 0 for 37 titles
 - 1 % for 3 titles
 - 2 % for 1 title

- MONTHLY MAGAZINES
concerning the 74 publications studied , the discrepancies
between the audiences measured by the two files are :
 - 0 for 4 titles
 - 1 % for 33 titles
 - 2 % for 29 titles
 - 3 % for 5 titles
 - 4 % for 3 titles

The percentage basis is the audience measured in the original file.

NEW EXPERIENCES ON MERGING IN ITALY

TABLE 2

COMPARISONS

between

ISPIPRESS

ISEGIPRESS

after fusion

- 8 WEEKLY MAGAZINE GROUPS
comparison through the CHI SQUARE method between PENETRATION RATES
before and after fusion crossed by the socio-demographic variables
(21 classes) :
 - 168 measures of which 0 are significant.

- 14 MONTHLY MAGAZINE GROUPS
comparison through the CHI SQUARE method between PENETRATION RATES
before and after fusion crossed by 5 socio-demographic variables
(21 classes) :
 - 294 measures of which 0 are significant.

NEW EXPERIENCES ON MERGING IN ITALY

TABLE 3

COMPARISONS

between

ISPIPRESS

ISEGIPRESS

after fusion

- 8 WEEKLY MAGAZINE GROUPS
comparison through the CHI SQUARE method between COMPOSITION RATES
before and after fusion crossed by 5 socio-demographic variables
(21 classes) :
 - 168 measures of which 0 are significant.

- 14 MONTHLY MAGAZINE GROUPS
comparison through the CHI SQUARE method between COMPOSITION RATES
before and after fusion crossed by 5 socio-demographic variables
(21 classes) :
 - 294 measures of which 0 are significant .

NEW EXPERIENCES ON MERGING IN ITALY

TABLE 4

COMPARISONS

between

ISPIPRESS

ISEGIPRESS

after fusion

- 8 WEEKLY MAGAZINE GROUPS
comparison through the CHI SQUARE method between the audience
cross-tabulated itself :
 - 64 measures of which 0 are significant.

- 144 MONTHLY MAGAZINE GROUPS
comparison through the CHI SQUARE method between the audience
cross-tabulated with itself :
 - 196 measures of which 0 are significant.

TABLE 5

COMPARISONS

between

ISPIPRESS

ISEGIPRESS

after fusion

- WEEKLY MAGAZINES
comparison through the CHI SQUARE method between the audience for 41 titles cross-tabulated with that of 8 groups of weekly publications :
 - 328 measures of which 14.6 % are significant.

- MONTHLY MAGAZINES
comparison through the CHI SQUARE method between the audience for 74 titles cross-tabulated with that of 14 groups of monthly publications :
 - 1036 measures of which 5.7 % are significant.

NEW EXPERIENCES ON MERGING IN ITALY

4) EXPERIMENTAL MEDIA PRESS/PRODUCTS SURVEY

CHARACTERISTICS OF THE ITALIAN MEDIA PRESS/PRODUCTS SURVEY

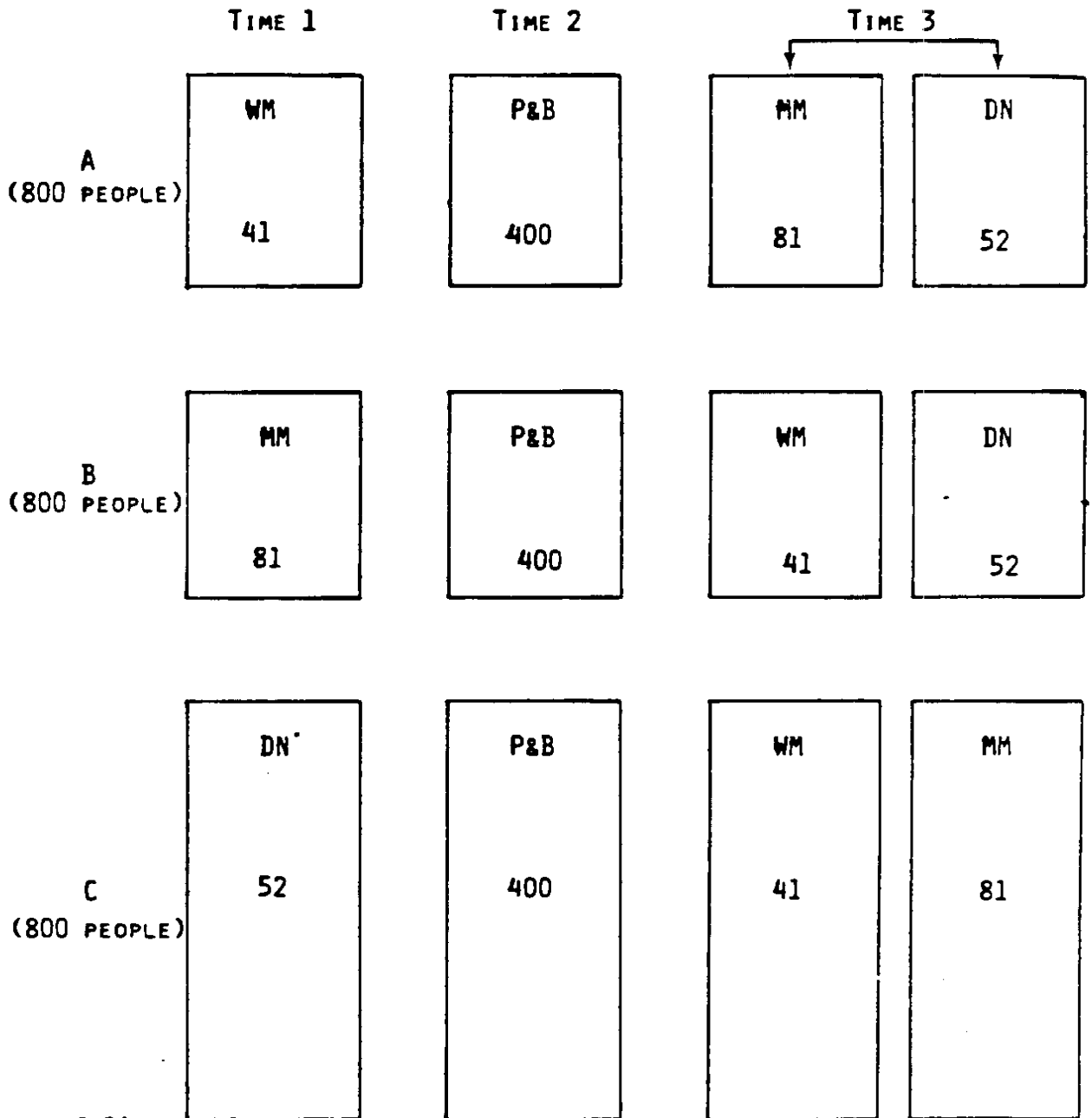
The basic design of the new survey is as follows (see also figure 5):

<u>Sub-samples</u>	<u>information collected</u>		
	compilation time 1	compilation time 2	compilation time 3
	(self-administrated)		
A	weekly magazines	products and brands	monthly magazines and daily newspapers
B	monthly magazines	products and brands	weekly magazines and daily newspapers
C	daily newspapers	products and brands	weekly magazines and monthly magazines

NEW EXPERIENCES ON MERGING IN ITALY

FIGURE 5
EXPERIMENTAL SURVEY

SUB-SAMPLES



LEGENDA:

- W M = WEELKY MAGAZINES
- P & B = PRODUCTS AND BRANDS
- M M = MONTHLY MAGAZINES
- D N = DAILY NEWS PAPERS

NEW EXPERIENCES ON MERGING IN ITALY

We used the same sample design, applied to traditional media-press surveys , also for the sub-samples A,B,C , even though the sample size of C was double of that of A and B.

The reason for different sample sizes is twofold: we have to sufficiently represent the local newspapers reality and account for over-sampling.

We have to resolve two problems categories, namely :

(A) first category

- missing data coming from time 2 failures
- missing data coming from time 3 failures

(B) second category

- comparison between audiences measured from times (compilation times) and eventually following treatment to make audiences statistically equivalent.

Interventions required by problem categories A and B have to be multiplied by the 4 groups of information :

- weekly magazines
- monthly magazines
- daily newspapers
- products and brands

that will be treated separately and independently.

The work hypothesis are as follows :

- the correct audience value for magazines and newspapers could be measured only using answers collected at compilation time 1, although we believe in the statistical equivalence of audiences evaluated on samples interviewed at different compilation times
- the reader profiles are consistent if measured on the total sample A + B + C, where each target is well represented.

NEW EXPERIENCES ON MERGING IN ITALY

MEDIA-PRESS/PRODUCTS DATA TREATMENT

Using the part of fusion theory concerning similarity concepts and related merging rules , we have developed two new processes :

- (1) EXTENTION of information for questionnaires with missing data (see figure 6)
- (2) audience CALIBRATION working for each group of media-press information over the total sample A + B + C (see figure 7).

The complete data treatment, after data cleaning, is so structured :

- (1) 3 independent, common design , WEIGHTING procedures, working on 19 geographic areas and controlling 7 variables at the same time
- (2) ANTE ANALYSIS concerning study of original estimates obtained from different samples
- (3) EXTENTION
- (4) CALIBRATION
- (5) POST ANALYSIS concerning study of results calculated after data treatment phases 3 and 4.

CONCLUSIONS AND PERSPECTIVES

At present it is not possible to form any conclusion about the result of data treatment procedures we are developing and testing.

In one month we will have the final results of the experiment and we will be able to make a validity judgement about how the new technics work, especially concerning

- difficulty to complete missing data
- suitability of profile control and correction
- extension or compression of audiences, respecting the audiences figures obtained from primary samples.

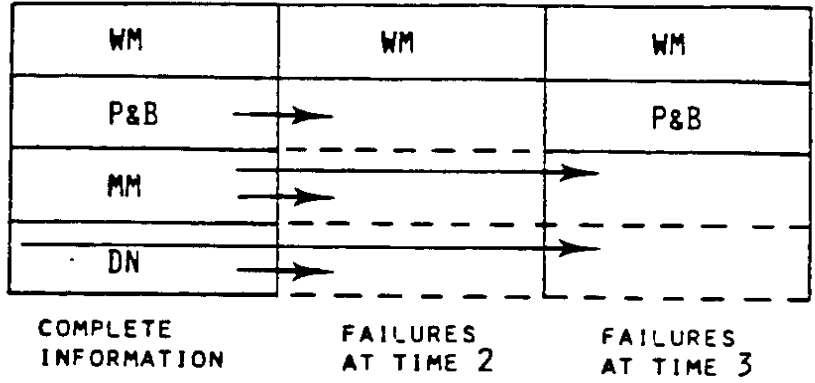
If the experiment is successful, we will extend the same methodology to a complete survey of more than 25,000 people for each wave, taking place two times per year.

FIGURE 6
EXPERIMENTAL SURVEY

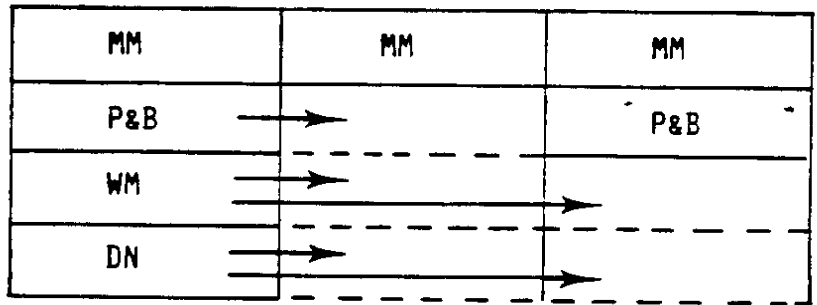
EXTENSION (HORIZONTAL PROCEDURES)

SUB-SAMPLES

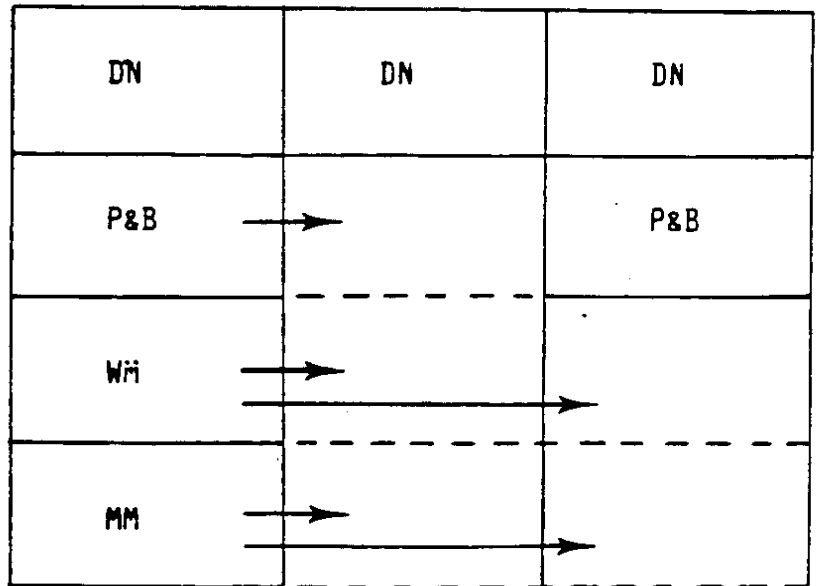
A



B



C

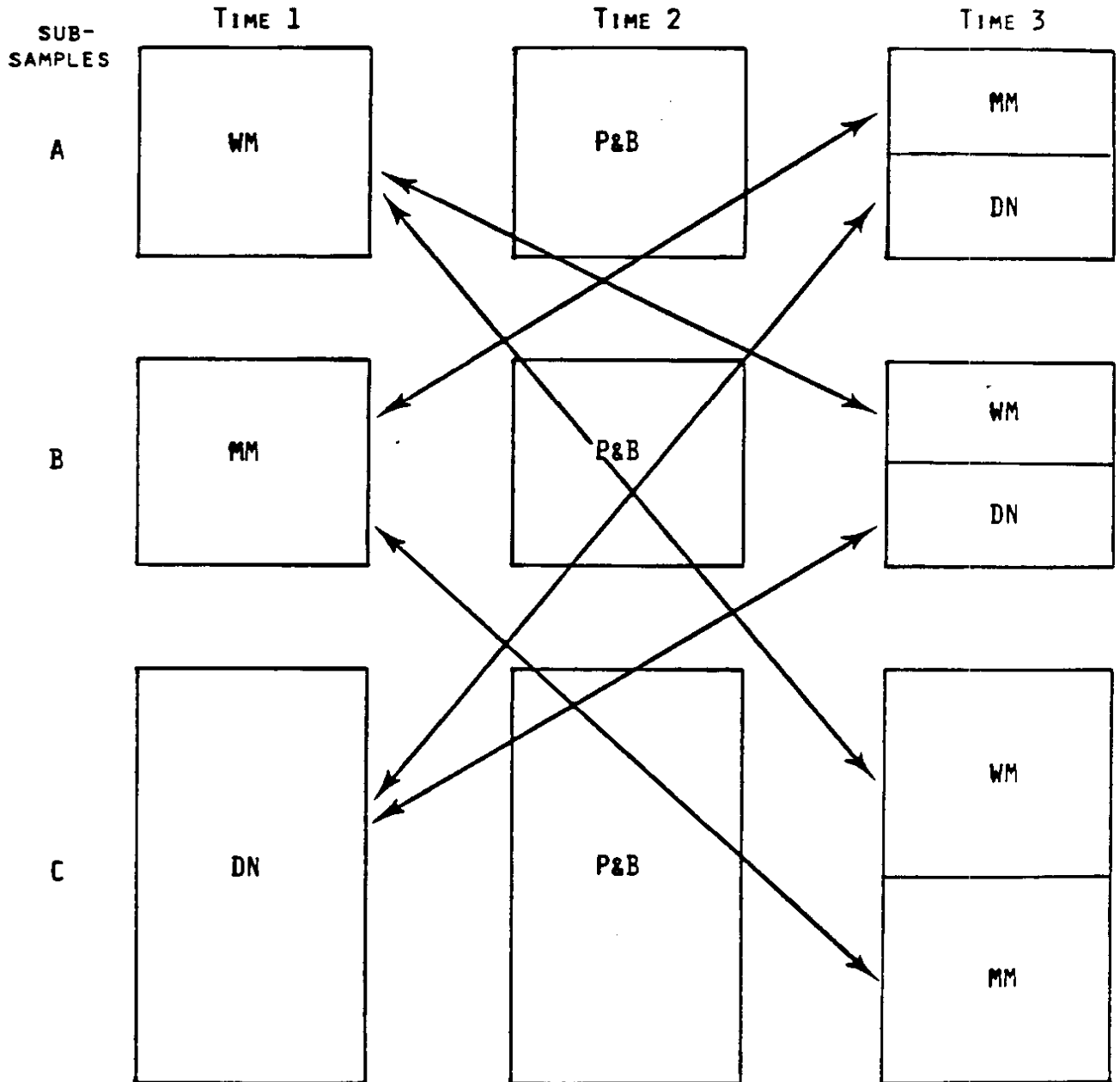


LEGENDA:

- W M = WEEKLY MAGAZINES
- P & B = PRODUCTS AND BRANDS
- M M = MONTHLY MAGAZINES
- D N = DAILY NEWS PAPERS

FIGURE 7
EXPERIMENTAL SURVEY

CALIBRATION (DIAGONAL PROCEDURES)



GENERAL CONCLUSIONS

The recent establishment of AUDIPRESS has promoted some new initiatives with the intent to measure variables not present in the traditional media-press surveys.

AUDIPRESS is now very concentrated on measurement of publication middle issue life cycles during a pre-fixed period, through modelling technics that can allow a post-intervention on the traditional data.

We can conclude by saying that the main directions we follow at the moment are three and they are strongly connected

- traditional media-press surveys (fusion processes)
- experimental media-press/products and brands surveys (extension and calibration technics)
- middle issue life cycles studies (mathematical models).