**Wally Langschmidt**
Consultant, South Africa

# 9.1 THE MAJOR REASON WHY OUR READERS-PER-COPY FIGURES ARE TOO HIGH

*Editorial note: the very lengthy paper by Wally Langschmidt distributed at Barcelona had in fact been written for a different purpose. The paper he actually delivered to the Symposium appears below.*

We were lucky in South Africa that the 1980 census included a number of durable items. During the same period our large All Media and Products Survey (AMPS) included the same durables. Table 1 provides a comparison between the census and the AMPS results. It is clear from this table that the results of the survey are remarkably close to those of the census, the correlation coefficient being .998 and the average difference about 3% – only 0.6% on the total universe.

However, if we compare the 'claimed circulation' at this time with the official ABC figures for all publications included in the survey then we find a difference of 96%. The sixty-four thousand dollar question is 'why'?

**Table 1**

**The accuracy of AMPS**

| Item | 1980 Census '000 | 1980 AMPS '000 |
|---|---|---|
| Cars/minibuses | 4,008 | 3,801 |
| Telephone ('Pvt') | 1,798 | 1,823 |
| Refrigerators | 1,286 | 1,246 |
| Electrical stoves | 1,287 | 1,131 |
| TV sets | 1,117 | 1,102 |
| Washing machines | 1,018 | 1,019 |
| Motor bikes | 344 | 338 |
| Swimming pools | 146 | 132 |
| Caravans | 135 | 140 |
| Average per item | 1,238 | 1,192 |

During the past 16 years I have examined 56 factors that can and do play a part in readership measurement. They cannot account for an apparent 'inflation' of 96%: there must be a further missing factor.

At the Salzburg Symposium in 1985 one delegate put forward a modified version of an old dictum: 'a good theory with a good model and a sound hypothesis is half the problem solved'. The basic hypothesis put forward became my 57th factor: 'is there a theoretical difference between the results to be expected for a 'single' sample compared with a series of matched samples spread evenly over time?' There was not time or money to test the hypothesis in real life, so I decided to 'test' it via a series of 'Theoretically Simulated Surveys' (TSS).

Let us start off with a bit of theory. The number of reading combinations that are possible with a definite number of issues of a specific publication can be calculated mathematically. For example, one issue has only two reading possibilities – yes and no. Two issues have four possibilities, three have eight, and so on, as shown in Figure 1. The number of possible combinations is $2^n$ where n equals the number of issues in the scale. With six issues in our scale – the number we actually use in AMPS – the possible number of reading combinations is 64.

In Figure 2 we show how the 64 possible combinations are derived. In our readership studies over the past 40 years we have found that the replies in the low frequency groups are less reliable than in the higher frequency groups.

## 9.1 THE MAJOR REASON WHY OUR READER-PER-COPY FIGURES ARE TOO HIGH

**Figure 1**

**The basic concepts used in the TSS (Theoretically Simulated Surveys)**

The number of possible reading combinations

| 1 issue | Yes | No |
|---|---|---|
| 2 issues | Yes | Yes |
| | Yes | No |
| | No | Yes |
| | No | No |

| 3 issues | | *Issue Number* | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| 3 out of 3 | Yes | Yes | Yes | Yes |
| 2 out of 3 | Yes | Yes | Yes | No |
| | | Yes | No | Yes |
| | | No | Yes | Yes |
| 1 out of 3 | Yes | Yes | No | No |
| | | No | Yes | No |
| | | No | No | Yes |
| 0 out of 3 | Yes | No | No | No |

1 issue = $2^1$ = 2 3 issues = $2^3$ = 8 6 issues = $2^6$ = 64*
2 issues = $2^2$ = 4 4 issues = $2^4$ = 16 12 issues = $2^{12}$ = 4096

For the Theoretically Simulated Surveys it was decided to make use of the 15 possible 'readers' in the two out of six frequency group. We also assume that we are dealing with a monthly magazine, that each survey is in fact a census of all 15 readers in the group, that we undertake our census surveys on the last day of the month, and that we are using the Through-the-Book method for both the frequency and recency questions. I must stress that (in this TSS) universe, census, model and sample survey all mean or represent the same thing.

In Figure 3 we look at all 15 possible combinations for reading two out of six issues between January and June. In this model it can be seen, for example, that reader number 1 read the January and February issue, reader number 2 the February and March issues, and so on. Viewing the full six issues from January to June it can be seen that each issue was read by five out of the 15 total readers in the group, providing us with the correct issue-probability of .3333, and that each reader in the group had

**Figure 2**

**Possible reading combinations with 6 issues**

Maximum possible combinations =

$$ {}^{n}Cr = \frac{n!}{(n - r)!\, r!} = \frac{6!}{(6 - r)!\, r!} = \frac{720}{(6 - r)!\, r!} $$

Number of possible reading combinations

| *Combination* | *6 Cr* | *No. of possible combinations* |
|---|---|---|
| 0 out of 6 (r = 0) | 720 ÷ 720 | = 1 |
| 1 out of 6 (r = 1) | 720 ÷ 120 | = 6 |
| 2 out of 6 (r = 2) | 720 ÷ 48 | = 15 |
| 3 out of 6 (r = 3) | 720 ÷ 36 | = 20 |
| 4 out of 6 (r = 4) | 720 ÷ 48 | = 15 |
| 5 out of 6 (r = 5) | 720 ÷ 120 | = 6 |
| 6 out of 6 (r = 6) | 720 ÷ 720 | = 1 |

read two out of the six issues, giving us the correct reader-probability of .3333.

Let us now take a quick look at the 'results' of our Theoretically Simulated Surveys. Figure 4 summarises the TSS results for the two out of six group. It is clear from this that each issue received its correct figure of five issue readers out of the group total of 15 – ie 33.3% (Line A). Line B records the cumulative issue readership, Line C shows how we are adding new readers to our results as we cover additional issues, and Line D accumulates the number of new readers added per issue surveyed. Line E shows how the number of issues read by each recorded reader climbs from one for the first issue to two for the sixth issue. The important point to stress is that we only reach the correct reader-probability after all six issues have been covered.

The results shown in tabular form in Figure 4 are shown in Figure 5 in graphical form.

**Figure 3**

**Simulated model for the 2 out of 6 group**

| *Reader Details* | *Month* J | F | M | A | M | J | *Total* |
|---|---|---|---|---|---|---|---|
| 1 | ● | ● | - | - | - | - | 2 |
| 2 | - | ● | ● | - | - | - | 2 |
| 3 | - | - | ● | ● | - | - | 2 |
| 4 | - | - | - | ● | ● | - | 2 |
| 5 | - | - | - | - | ● | ● | 2 |
| 6 | ● | - | ● | - | - | - | 2 |
| 7 | - | ● | - | ● | - | - | 2 |
| 8 | - | - | ● | - | ● | - | 2 |
| 9 | - | - | - | ● | - | ● | 2 |
| 10 | ● | - | - | ● | - | - | 2 |
| 11 | - | ● | - | - | ● | - | 2 |
| 12 | - | - | ● | - | - | ● | 2 |
| 13 | ● | - | - | - | ● | - | 2 |
| 14 | - | ● | - | - | - | ● | 2 |
| 15 | ● | - | - | - | - | ● | 2 |
| Issue rdrs | 5 | 5 | 5 | 5 | 5 | 5 | 30 |
| 'New' rdrs | 5 | 4 | 3 | 2 | 1 | - | 15 |
| Cume rdrs | 5 | 9 | 12 | 14 | 15 | 15 | 15 |

**Figure 4**

**Summary of the results of the 6 simulated monthly surveys**

| *Details* 2 out 6 | *Issues covered* Jan | Feb | Mar | Apr | May | Jun |
|---|---|---|---|---|---|---|
| A. Readership of each new issue | 5 | 5 | 5 | 5 | 5 | 5 |
| B. Cumulative issue readership | 5 | 10 | 15 | 20 | 25 | 30 |
| C. No of 'new' readers added per issue | 5 | 4 | 3 | 2 | 1 | - |
| D. Cumulative readers | 5 | 9 | 12 | 14 | 15 | 15 |
| E. No of issues recorded per actual reader ( that is B ./. D ) | 1.00 | 1.11 | 1.25 | 1.43 | 1.67 | 2.00 |

## 9.1 THE MAJOR REASON WHY OUR READER-PER-COPY FIGURES ARE TOO HIGH

The graph clearly shows how issue-readership grows linearly from five to 30 over the six issues *but that actual readers do not.* In other words, if we want to credit all the readers in a specific group with the 'correct' number of issues we have to consider six issues at a time!

Figure 6 shows that if we want to 'credit' our sample universe of 15 readers with 30 'copies' over the previous six issue-periods we in fact have to 'go back' to August of the previous year. The figure also shows that we undertake our six successive monthly surveys, with each survey representing our full universe of 15 readers, then out of the 36 months (6 x 6) covered in the Simulated Surveys 15 'months' representing 75 issue-readers ('copies') originated from outside our January to June fieldwork period.

**Figure 5**

**The 'build-up' in issue-readers vs number of actual readers over 6 issue-periods**

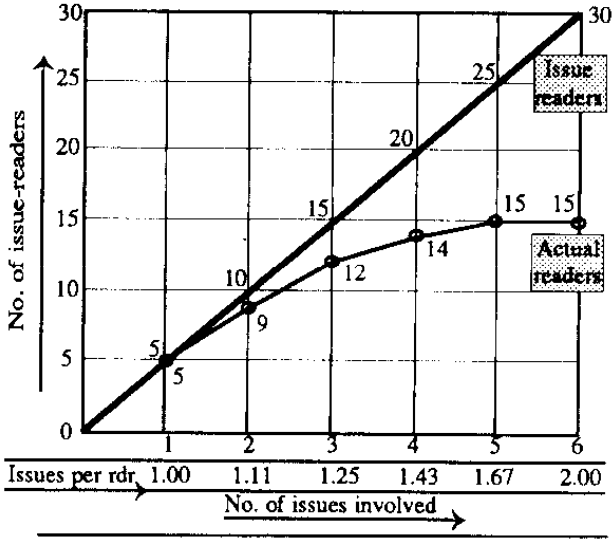



**Figure 6**

**Theoretical copies 'consumed' in 6 TSS**

| | | J | F | M | A | M | J | T | M | |
|---|---|---|---|---|---|---|---|---|---|---|
| Issues | A | 5 | | | | | | 5 | 1 | |
| | S | 5 | 5 | | | | | 10 | 2 | |
| | O | 5 | 5 | 5 | | | | 15 | 3 | 15 months 75 'copies' |
| | N | 5 | 5 | 5 | 5 | | | 20 | 4 | |
| | D | 5 | 5 | 5 | 5 | 5 | | 25 | 5 | |
| 'Current' 6 month period | J | 5 | 5 | 5 | 5 | 5 | 5 | 30 | 6 | |
| | F | | 5 | 5 | 5 | 5 | 5 | 25 | 5 | |
| | M | | | 5 | 5 | 5 | 5 | 20 | 4 | 21 months 105 'copies' |
| | A | | | | 5 | 5 | 5 | 15 | 3 | |
| | M | | | | | 5 | 5 | 10 | 2 | |
| | J | | | | | | 5 | 5 | 1 | |
| | T | 30 | 30 | 30 | 30 | 30 | 30 | 180 | 36 | months |

## 9.1 THE MAJOR REASON WHY OUR READER-PER-COPY FIGURES ARE TOO HIGH

**Figure 7**

**Reporting pattern on the 6-month ABC period for the 2 out of 6 group**

| Period | Monthly sample | Cume sample | Number of 'current' issues seen | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Jan | 15 | 15 | 10 | 5 | - | | | | |
| Feb | 15 | 30 | 6 | 8 | 1 | | | | |
| Mar | 15 | 45 | 3 | 9 | 3 | | | | |
| Apr | 15 | 60 | 1 | 8 | 6 | | | | |
| May | 15 | 75 | - | 5 | 10 | | | | |
| Jun | 15 | 90 | - | - | 15 | | | | |
| Totals | 90 | 90 | 20 | 35 | 35 | | | | |
| Claims grossed up | | | - | 35 | 70 | | | | |

With 105 issue-readers out of the assumed total of 180 (6 x 30) the yield is a 'current' 6-period proportion of .5833 – ie, 105 divided by 180.

This ratio of .5833 'in' the 'current' 36 periods remains constant whether we are dealing with monthlies, fortnightlies, weeklies or dailies.

In Figure 7 we look at the results of the same six successive theoretical surveys if we showed each informant the six most recent issues but only 'accept' issues from the current six month period (January to June) as being 'permissible'. Out of the cumulative sample of 90 'informants' over the period January to June we will have:

20 reporting 'nil' 'current' issues = 0 'copies'
35 reporting 1 'current' issues = 35 'copies'
35 reporting 2 'current' issues = 70 'copies'
Total = 105 'copies'

But our 'assumed' figure is 180 'copies' (90 x .333 x 6), and this yields the same 'step-down' factor of .5833 (105 divided by 180).

The crux of the matter is that issue-probability and reader probability are *not* synonymous. Correct issue-probability can be achieved after one issue, but correct reader-probability requires six issues, and only 21 of the 36 issues required fall in the 'current' six period (and 21 divided by 36 is .5833). The point is further developed in Figure 8. The real position is that we have three probabilities to contend with – issue probability, personal probability, and group probability.

Before we take a closer look at these three concepts we must consider two statistical laws:

**(a) The addition law.** "The probability that an event will occur in one of several possible ways is calculated as the sum of the probabilities of

**Figure 8**

**The 1 out of 6 frequency group and 'when last' readership**

| | Issue Number | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| Periods ago | 6 | 5 | 4 | 3 | 2 | 1 | |
| Read 1 out of 6 | 1 | 1 | 1 | 1 | 1 | 1 | 6 |

▲ Fieldwork here

**The average time lapse** over the six issues for this group is therefore ...

$(1 + 2 + 3 + 4 + 5 + 6) \,./.\, 6 = 3.5$ **Periods** →

If $P = \frac{1}{TL}$ we have $\frac{1}{3.5} = .2857$ and not .1667 →

**But**: Multiply .2857 by the factor .5833 and we are back to the truth .1667

the occurrence of the several different possible ways" (Moroney 1965).

**(b) The multiplication law.** "If two outcomes are independent of each other, the probability of both outcomes occurring is the product of the probabilities of each of the outcomes" (O'Muircheartaigh & Francis 1981).

Let us now look at 'six equal issues' of a publication. In virtually all readership models we assume that our average issues are equal as far as audience size and circulation are concerned. If our universe of publication contains six issues, the p for one issue is one-sixth, or .1667, for two issues two sixths, or .3333, and for six issues six-sixths, or 1.0000. We also know, from the combination theory, that there are only six ways in which one issue out of six can be read. The personal or 'reader' probability for a single issue is therefore also one-sixth, or .1667.

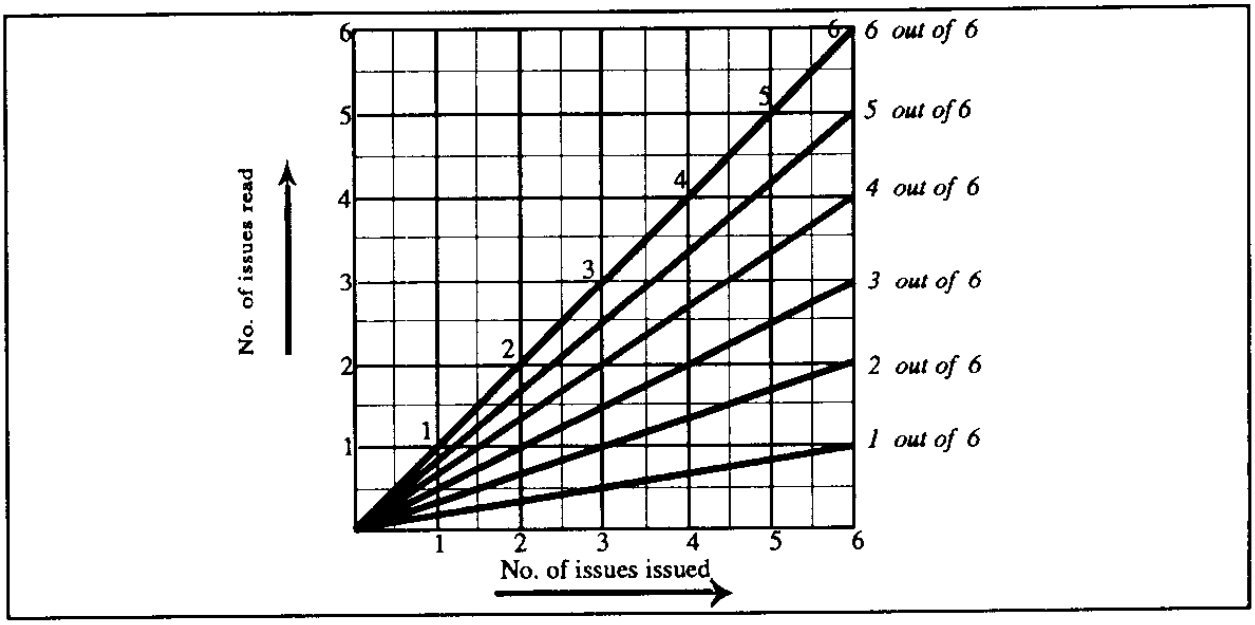
If we now combine issue probability of one-sixth with the reader probability of one-sixth the law of multiplication must apply, and the combined probability of issues and readers becomes one-sixth of one-sixth, or one-thirty-sixth – .0276.

Let us now, at the group level, look at the relationship between the number of issues issued and the number of issues read by the average reader in each group. At the group level the number of issues read by all the

## 9.1 THE MAJOR REASON WHY OUR READER-PER-COPY FIGURES ARE TOO HIGH

**Figure 9**

**Number of issues read at the group level**




readers in each frequency group grows linearly in that group. This is clearly shown in Figure 9. If we look at the maximum-probability group (that is, the six out of six group) we cannot credit them with having read more than one issue if we are only considering one issue. Maximum probability is one out of one, two out of two and three out of three etc. We can thus credit this group with six issues only after six issues have been issued.

Figures 10 and 11 should be considered together: Figure 10 is in fact a tabular version of the other. In this table it can clearly be seen how, at the group level, the law of addition applies both horizontally and vertically. However, if we combine reader probability then the law of multiplication must apply. This fact is reflected in the probabilities shown along the diagonal, where the two 'equal' probabilities of 'issues' and 'group readers' meet or coincide.

Figure 11 is an interesting one. The six out of six diagonal line shows the probabilities we *assume* to be correct for the various frequency groups. The 'bottom' line, however, reflects the true reader probabilities within each group. Thus, for example, readers in the three out of six group will, as a group, build up to three issues over six issue periods, but the individual readers in the group cannot be 'credited' with more than the three issues they claim to have read. The bottom line, in fact, reflects the actual probabilities one would expect via the Through-the-Book method of readership reporting.

Figure 12 is a purely theoretical version of Figure 11, and shows how the combining of issues and readers 'reduces' the dual probability of the combination.

Let us now take a look at a 'real life' fieldwork situation, illustrated in Figures 13 and 14. As

**Figure 10**

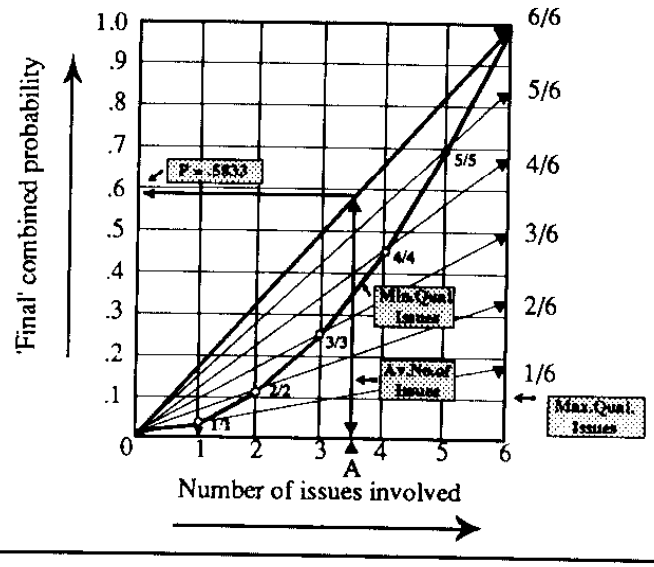
**Growth patterns in group probabilities as the number of issues, out of six, that are included is increased**

| Frequency of reading group | *Number of issues included* | | | | | | Group probability |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1 out of 6 | .0278 | .0556 | .0833 | .1111 | .1389 | .1667 | 1/6 |
| 2 out of 6 | .0556 | .1111 | .1667 | .2222 | .2778 | .3333 | 2/6 |
| 3 out of 6 | .0833 | .1667 | .2500 | .3333 | .4167 | .5000 | 3/6 |
| 4 out of 6 | .1111 | .2222 | .3333 | .4444 | .5556 | .6667 | 4/6 |
| 5 out of 6 | .1389 | .2778 | .4167 | .5556 | .6944 | .8333 | 5/6 |
| 6 out of 6 | .1667 | .3333 | .5000 | .6667 | .8333 | 1.0000 | 6/6 |
| Issue prob. | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 | 6/6 | 1.00 |

Note: Horizontally' and 'vertically' the **Law of Addition** applies
Combining *Readers and Issues* the **Multiplication Law** must apply (diagonally)

**Figure 11**

**Combined issue and personal probabilities**



# THE MAJOR REASON WHY OUR READER-PER-COPY FIGURES ARE TOO HIGH

we are doing our fieldwork we are moving across calendar time in units of issue periods To achieve correct reader probabilities we have to consider six issues at a time, and the maximum effect that this six issue measurement method can have on reader probabilities is shown for the six out of six group in Figure 15.

If we have multiple fieldwork cycles we must be including multiple issues in our measurement method. I sincerely believe that all of us have been confusing issue probability with personal or reader probability. As we move across time we must be including different issues whether we like it or not.

In Figure 15 I have attempted to show results from the Theoretically Simulated Surveys, using a six issue scale and then comparing the 'true' cumes with the 'assumed' cumes and the six issue cumes. Here again we come up with the deflation factor of .5833 after six issues. At

**Figure 12**

**'Single' vs 'dual' probability**

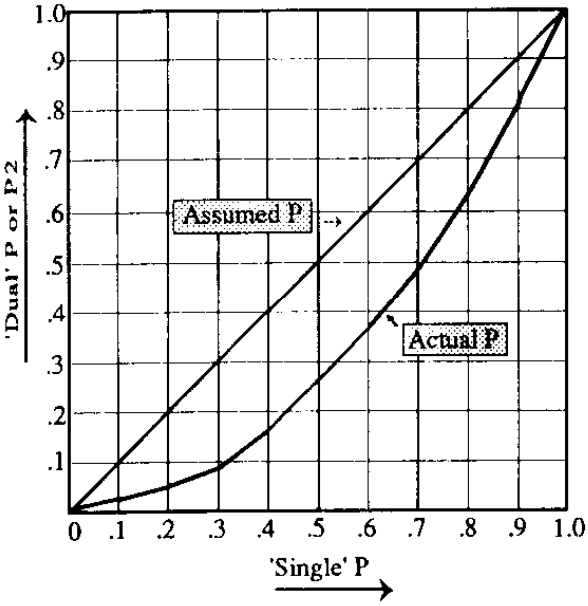



**Figure 13**

**The position up to here**

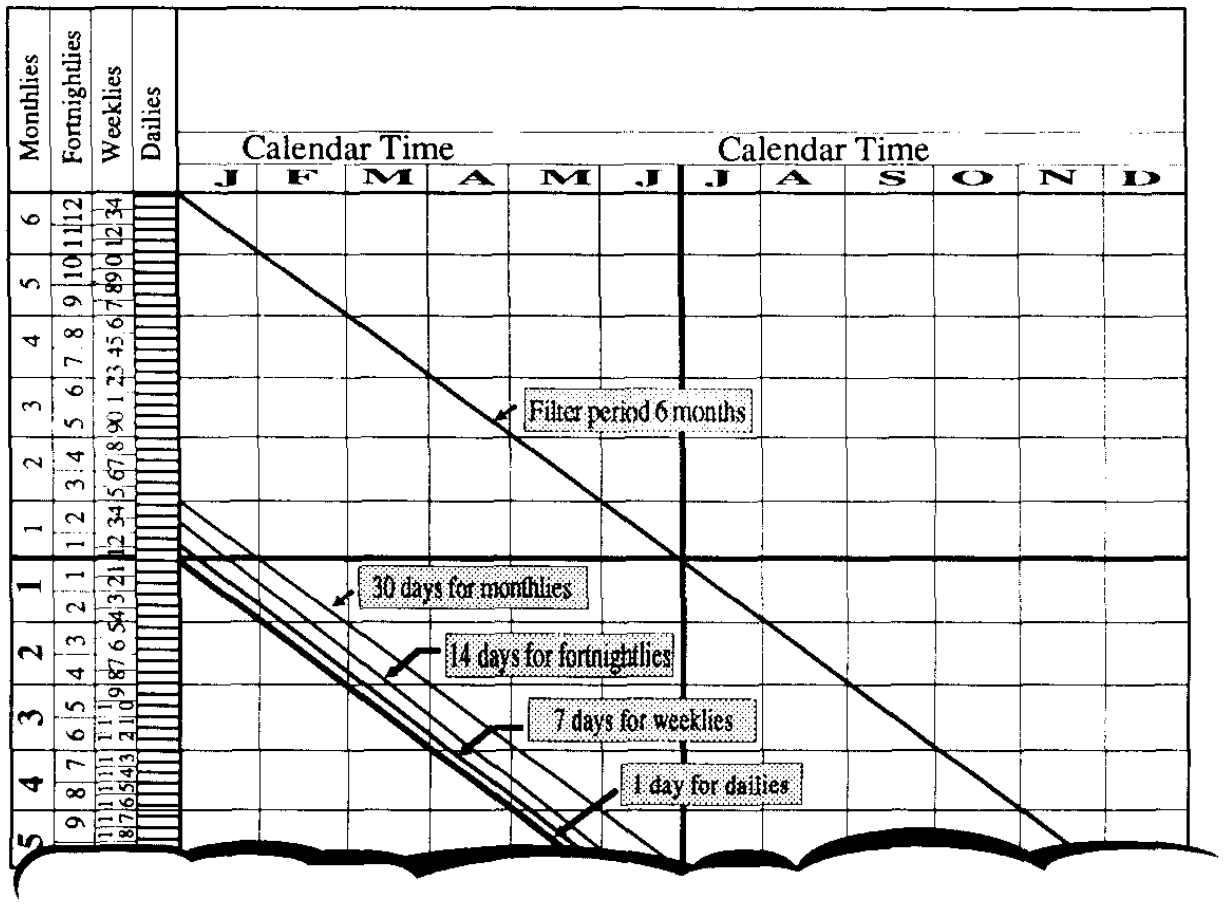
As we are doing our fieldwork we are moving across **Calendar-Time** in units of **Issue-Periods**

*The Time-Scale Scene*

| Issue Periods | Publication Time | Calendar Time | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | J | | F | | M | | A | | M | | J | |
| | Monthlies | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
| | Fortnightlies | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | Weeklies | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 |
| | Dailies | | | | | | | | | | | | | | | | | | | | | | | | |

**Figure 14**

**The time-scale scene**




the group level the factor of .5833 appears to apply to all fieldwork cycles from one to six. If however we go beyond six cycles then the deflating factor required starts declining and if, as in the UK, there are more than six cycles of fieldwork, it will slowly decline until it reaches .7917 after 12 cycles. Figure 16 simply illustrates graphically the last column of Figure 15, and shows that even at the group level, and even after 12 fieldwork cycles, we still need a deflation factor of .7917. At the personal or individual reader level, however, the probabilities will follow the 'diagonal' or squared probabilities shown in Figure 10.

If the TSS logic is correct, than the proof of the pudding will be in the eating, and if actual survey buying claims are applied to TSS readership results than we should get TSS-deduced 'circulation' figures which are close to the official

## 9.1 THE MAJOR REASON WHY OUR READER-PER-COPY FIGURES ARE TOO HIGH

**Figure 15**

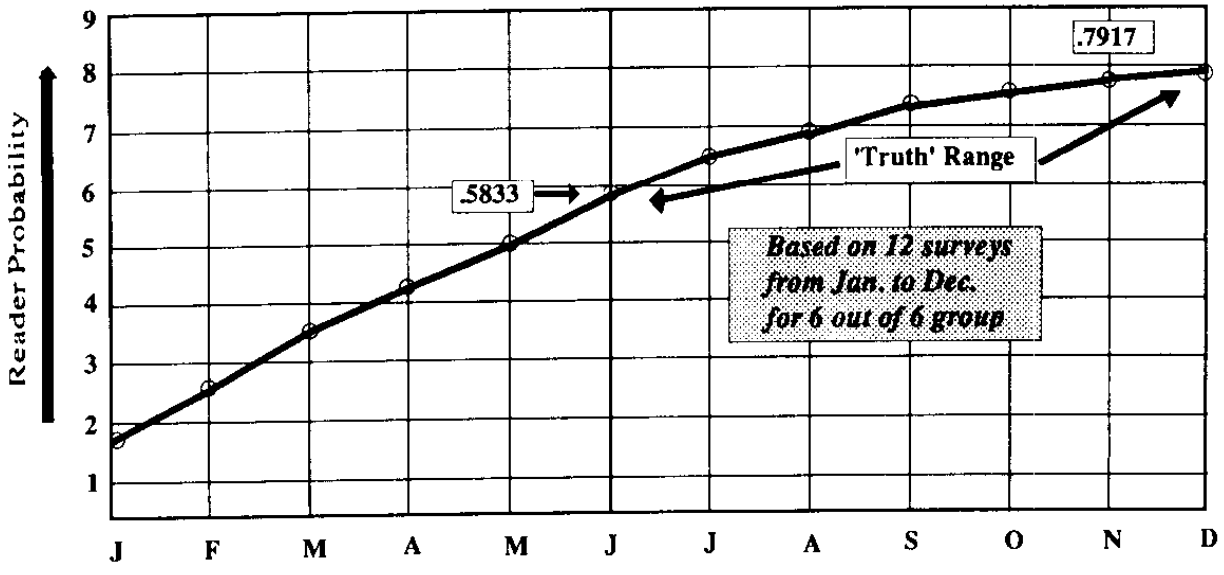
**Maximum readership growth pattern within the 6 out of 6 frequency group**

| A. Issue or survey number | B. True issue cumes | C. Maximum cumes with a 6pt. scale | D. 'Assumed' cumes with a 6pt. scale | E. 6pt. Scale 'deflation' factor (C./.D) |
|---|---|---|---|---|
| 1 | 1 | 1 | 6 | .1667 |
| 2 | 3 | 3 | 12 | .2500 |
| 3 | 6 | 6 | 18 | .3333 |
| 4 | 10 | 10 | 24 | .4167 |
| 5 | 15 | 15 | 30 | .5000 |
| 6 | 21 | 21 | 36 | .5833 → |
| 7 | 28 | 27 | 42 | .6428 |
| 8 | 36 | 33 | 48 | .6875 |
| 9 | 45 | 39 | 54 | .7222 |
| 10 | 55 | 45 | 60 | .7500 |
| 11 | 66 | 51 | 66 | .7727 |
| 12 | 78 | 57 | 72 | .7917 → |

Truth Range

**Figure 16**

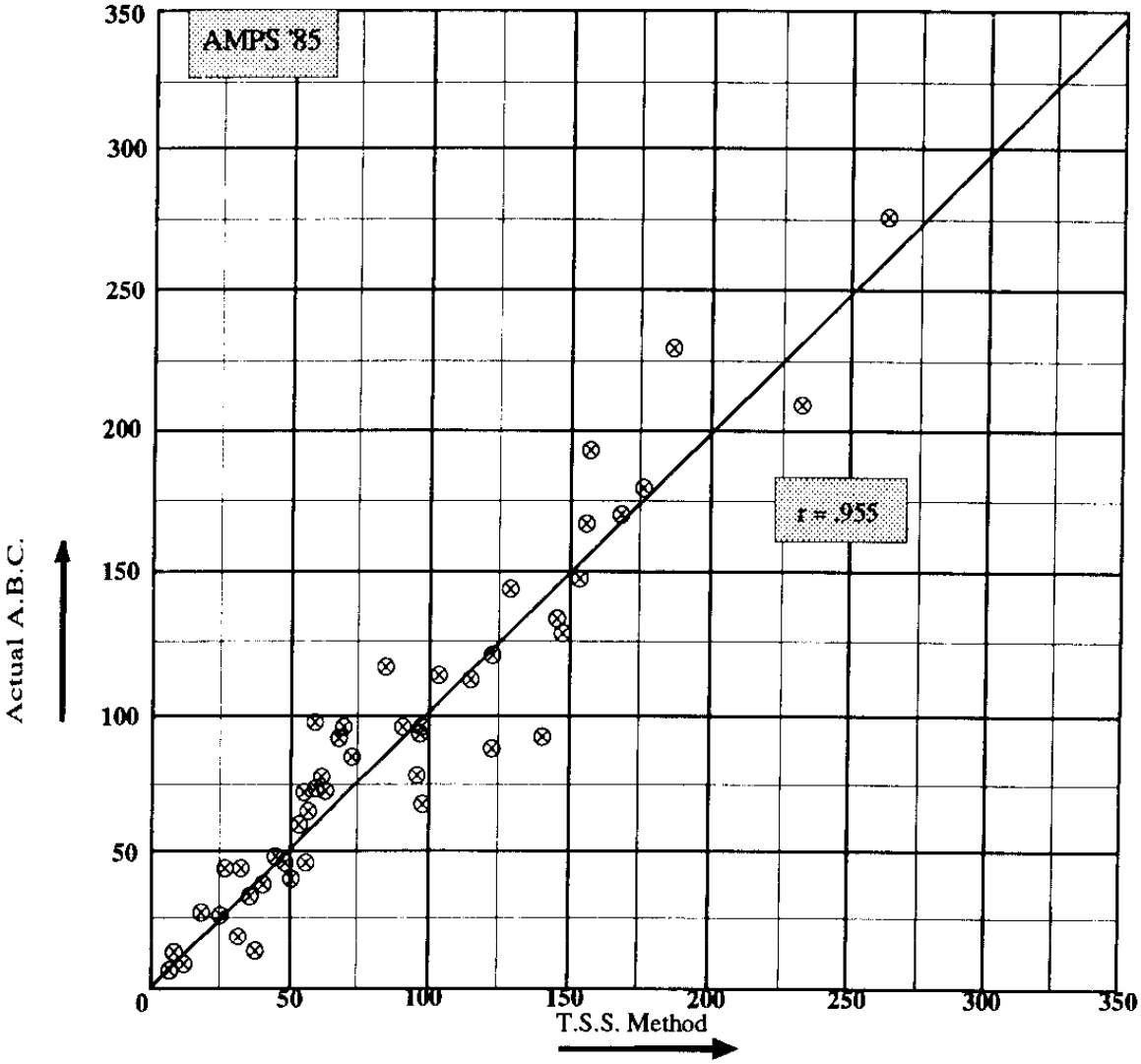
**Maximum growth pattern with 6-issue scale**



## 9.1 THE MAJOR REASON WHY OUR READER-PER-COPY FIGURES ARE TOO HIGH

**Figure 17**

**Actual vs TSS 'circulations'**

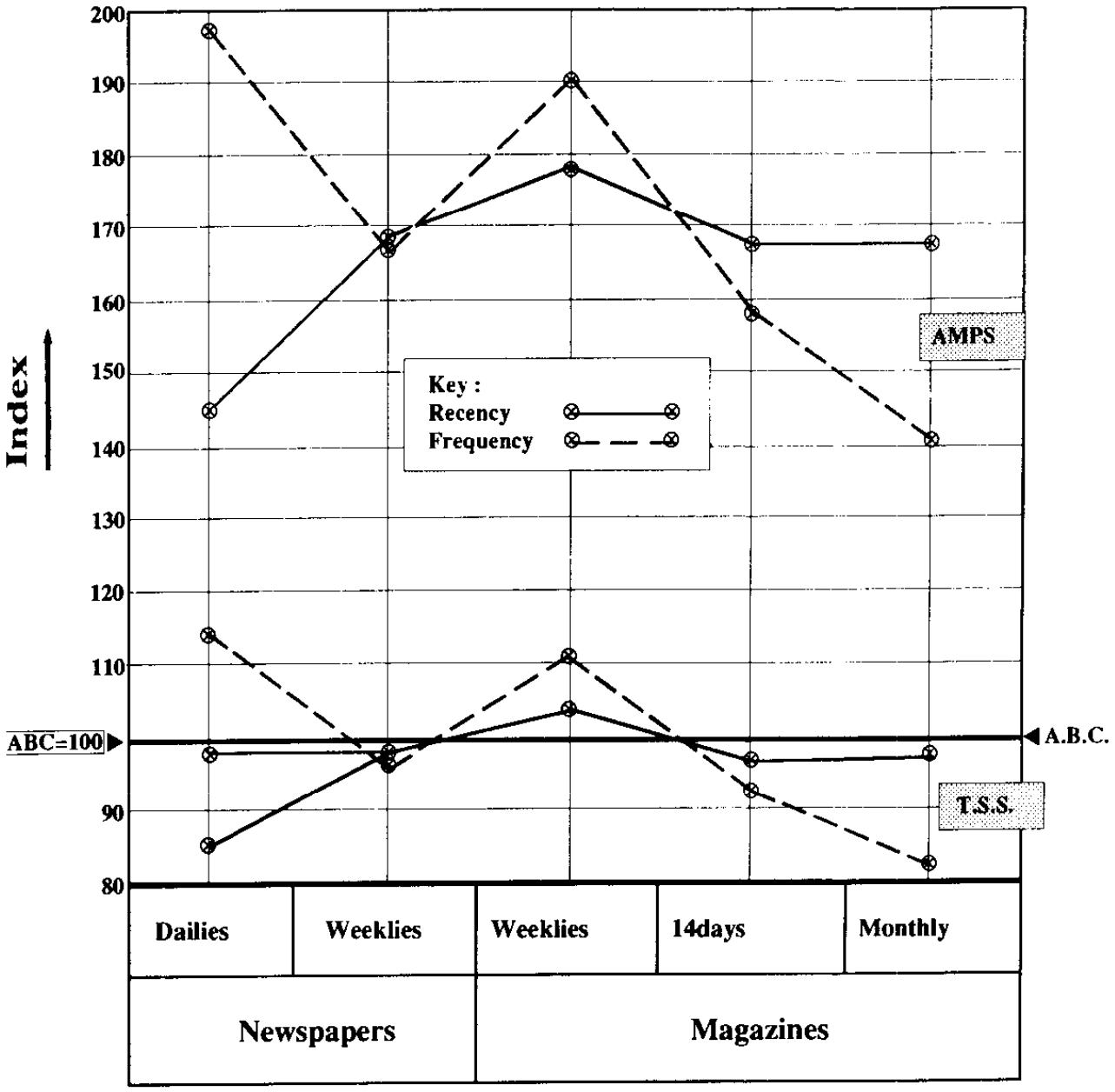



ABC figures. Figure 17 shows the relationship between the two: with a correlation of .955 the fit is quite good. In Figure 18 the various types of publication have been grouped together, and the survey-deduced circulations compared, on an index basis, with the official ABC figures (the ABC figures being 100).

The upper pair of lines in Figure 18 show the AMPS results via the recency and the frequency method. The lower pair show the same comparison but after the TSS logic has been applied to the results. The TSS results provide a good match with the ABC figures except in the case of the daily papers (and the Black

Figure 18

Actual vs survey 'circulations'



magazines, which are not shown here) where the index is 16% too low.

With daily papers having at least five to six times as many issues coming out as the other publication, and with, in the case of South Africa, (a) the bulk of the circulation of dailies originating in the large metropolitan areas, and (b) there being more fieldwork cycles in the metropolitan areas than elsewhere in the country, it is estimated that the 'deflation factor' in the case of daily papers is probably in the region of .688 (Figure 15 at the 8-issue level). Using the 8-issue level for the dailies the TSS Index climbs from 84 to 99 (from A to B).

The TSS method or approach reduces the readers-per-copy figures for newspapers from 6.3 to 3.7, and for the magazines from an average of 9.1 to 5.3.

The figures shown in Table 2 are taken from some tests undertaken by the ARF in the USA. I have commented earlier that the Through-the-Book method, being issue-specific, should theoretically be closer to the TSS figures. If the TSS deflation factor of .5833 is applied to the average recency results of the ARF and the Mediamark (ie, to 8.0) we obtain an index of 101 compared with the ARF Through-the-Book figure of 100.

I have travelled thousands of hours along the road to reliable readership, and have not yet reached the final destination. I leave it to younger researchers to take a closer look at the TSS and 'squared probability' concepts. But, finally, readership measurement is too complicated to claim that the TSS concept is a cure-all: there are 56 other factors that can and do play a part in measurement. I can myself shoot down some of my TSS comments, but the fact remains that the fantastic readers-per-copy inflation levels recorded via the recency method call for a drastic cure. I believe that the TSS logic had something in it, but one thing is certain – the road to reliable readership is always under construction.

**Table 2**

**ARF experiment**

| | A 67 magazine experiment | |
|---|---|---|
| | % | Index |
| (1) ARF – TTB | 4.6 | 100 |
| (2) ARF – RR | 8.3 | 180 |
| (3) Mediamark – RR | 7.7 | 167 |
| Average 2 + 3 | 8.0 | 174 |
| (4) TSS to average | 4.7 | 101 |

**References**

Moroney M (1965). *Facts from figures.* Penguin.

O'Muircheartaigh, Colm and Francis, DP (1981). *Dictionary of statistical terms.* Arrow Books.