Scott Cameron McDonald
Time Incorporated
New York, USA

4.7 PROCEDURE FOR THE USE OF SYNDICATED AUDIENCE RESEARCH TO DEVELOP SYNTHETIC COHORTS FOR HISTORICAL MEDIA ANALYSIS

WHAT IS COHORT ANALYSIS?

Cohort analysis has long been a powerful weapon in the analytical arsenal of demographers. For the most part, it has been used in the study of fertility; however in recent years, it has also been adapted to the study of diverse behavioural and attitudinal phenomena. Sociologists and economists have used it to study trends in female labour force participation, early retirement, juvenile delinquency, and the structure of economic opportunity. Political scientists have used cohort analysis to investigate trends in voter turnout, political party identification, and public opinion. Though the limitation of cohort analysis may ultimately prevent it from living up to the claims of its most enthusiastic supporters, it has nevertheless gained wide support in the social sciences as a useful device for studying the effects of human aging upon social, cultural and political change. Moreover, opportunities for cohort analysis have increased as repeated cross-sectional surveys - surveys that are particularly suitable to cohort analysis - have accumulated in data archives. In this paper, I propose the applicability of cohort analysis for the repeated cross-sectional time-series embedded in our syndicated audience research archives.

For those unfamiliar with the fundamental concepts of cohort analysis, a brief review may be helpful. The term *cohort* stems from the Latin term for the Roman military unit, and dictionaries still commonly define the word first to mean 'a group of warriors or soldiers'. In non-technical parlance, cohort refers to one's companion, associate, peer, or accomplice. The more technical usage favoured by demographers defines a cohort as some geographically or otherwise delineated population who shared a common significant life event within a given period of time.

Usually, when demographers speak of cohorts, they are referring to birth cohorts - that is, people born in a given year or set of years. Thus, when we speak of the Baby Boom Generation, we are only referring non-technically to the birth cohorts born in the United States between 1945 and 1964. However the term cohort can correctly be applied to other groups as well. Family sociologists might want to study divorce rates among different marriage cohorts - that is, people who were married for the first time in different years. Educational researchers might want to study occupational attainment among the Harvard classes of 1920, 1930, 1940, 1950, 1960, 1970 and 1980 - each of which represent distinct educational cohorts. Political analysts might want to follow the progress of the Freshman Congressmen of the 92nd Congress of the United States - a kind of political cohort. Though each of these examples refer to different types of cohorts, the term cohort nearly always refers to a birth cohort - and when it is used without any modifier, the implied modifier is almost always birth.

Cohort analysis, then, simply refers to any study that compares one or more cohorts on measures of some characteristic at two or more points in time. In effect, cohort analysis asks, "How are the people in Cohort X behaving now that they are 30 years old, compared with their behaviour when they were 20? How does their behaviour compare with that of people in Cohort Y when they were at the same ages?" The behaviour under examination could be working, voting, having babies, or, as I shall soon demonstrate, reading magazines.

One of the main objectives of cohort analysis is to help the researcher spot patterns of variation n the data that point to relative strengths of the age effects, cohort effects, or period effects. *Age effects* are those influences that stem directly from

aging; in theory, age effects exert the same influence on all people, regardless of their cohort membership or of the year of the survey. On the other hand, *cohort effects* derive from the unique experience of membership in a given cohort; for example, there is a considerable body of literature that suggests that Baby Boomers have distinct patterns of behaviour and consumption that set them apart from other generations; similarly, Glen Elder and his colleagues have shown that children raised during the Great Depression grew up to have generation-specific behaviour patterns. Technically speaking, these patterns are cohort effects. Finally, there are *period effects* - influences associated simply with a particular period of time, influences that affect all age groups and all cohorts in much the same way.

Before we consider some messy data from real life, it might be useful to examine tables that use hypothetical data to demonstrate 'pure' examples of these three effects. **Tables 1, 2** and **3** do just that, demonstrating 'pure' age effects, cohort effects, and period effects respectively. In **Table 1**, the level of the hypothetical variable rises regularly with each age step, regardless of cohort or year. In **Table 2**, each cohort is on its own specific track, and it maintains its unique level of the hypothetical variable, regardless of the year or of the cohort's age in that year. In **Table 3**, the level of the hypothetical variable is the same for all cohorts and all age groups, its value being determined only by the year of the survey. Since cohort analyses are usually presented graphically, these heuristic data are graphed, respectively, in **Figures 1** and **2** (**Table 3** cannot be clearly graphed in the absence of colour printing - Ed). As may be suspected, real data from actual surveys virtually never reveal patterns as clear and orderly as these, though tables and graphs showing patterns dominated by age effects are surprisingly frequent. Though one almost never sees a pure period effect or cohort effect, these effects often exert a visible influence on the patterns of variation. Unfortunately, there is no straightforward statistical way of disentangling these three theoretical effects, because two of the basic three effects will always be confounded with one another. That is, in any standard cross-sectional table that tabulates the percentage of variable X, by year, controlling for age, age and cohort effects are confounded in the columns, age and period effects are confounded in the cohort diagonals, and cohort and period effects are confounded in the rows. No statistical procedure can save us from this dilemma. Thus, cohort analysis merely refers to a method of organising and graphing these data from repeated cross sections so that, through visual inspection of the patterns of variation, one might assess the relative importance of the three effects.

TABLE 1
Cohort table in which all variation is due to age effects (hypothetical data)

	Year			
Age	*1940*	*1950*	*1960*	*1970*
20 - 29	40	40	40	40
30 - 39	45	45	45	45
40 - 49	50	50	50	50
50 - 59	55	55	55	55
60 - 69	60	60	60	60
70 - 79	65	65	65	65

WHY BOTHER? HOW CAN MEDIA RESEARCHERS USE COHORT ANALYSIS?

At this point, it is worth bringing the discussion down from these rather abstract planes to consider the possible applications of cohort analysis in media research. It is very unlikely that cohort analysis will help anyone to sell additional ad pages, nor is it likely to help circulators devise

TABLE 2
Cohort table in which all variation is due to cohort effects (hypothetical data)

	Year			
Age	*1940*	*1950*	*1960*	*1970*
20 - 29	50	40	30	20
30 - 39	60	50	40	30
40 - 49	70	60	50	40
50 - 59	80	70	60	50
60 - 69	90	80	70	60
70 - 79	100	90	80	70

TABLE 3
Cohort table in which all variation is due to period effects (hypothetical data)

	Year			
Age	*1940*	*1950*	*1960*	*1970*
20 - 29	70	60	50	40
30 - 39	70	60	50	40
40 - 49	70	60	50	40
50 - 59	70	60	50	40
60 - 69	70	60	50	40
70 - 79	70	60	50	40

marketing strategies. Cohort analysis certainly will not provide any clues to how to improve one's passalong figures, and I doubt that it will help resolve any of the tactical day-to-day decisions of business. Rather, its utility lies in its ability to clarify long-term historical trends in media audiences. In this regard, it can be of greatest use as a research tool for strategic planners and top-level executives who are charged with responsibility for the longer-range issues.

Consider, for example, the implications of the wide swings in the age structure of the United States population, swings that are lagged effects of the roller coaster postwar American fertility rates. As a result of these fluctuations, the 35-44 age group will swell over the next decade, while the 15-24 age group will diminish appreciably. If we were to assume that media tastes are shaped largely by age, and that succeeding generations age in more or less the same way - in other words, if we were to assume that age effects dominate and cohort effects are negligible - then we might predict that the aging of the gigantic postwar baby boom will create strong demand for business magazines, while depressing the demand for magazines that appeal more to young adults. Thus, to take examples from our own stable of magazines, we might, under these assumptions, expect bullish circulation estimates from *Fortune*, while settling for less ambitious ones from *Sports Illustrated*. In fact, we do not know that cohort effects are negligible, and in the case of *Sports Illustrated* there are reasons to think that the Baby Boomers are sustaining their interest in sports well beyond the ages that were characteristic of earlier generations. With a cohort analysis of 20 years of Simmons data, we were able to assess these issues empirically and develop growth forecasts based both upon our analysis of cohort-corrected historical trends and upon US Census population projections. For proprietary reasons, I shall not be sharing those analyses here, but I shall go through a different and interesting example involving *Time* and *Newsweek*.

An editor at *Time* once advanced the theory, perhaps based on a sampling of his own children or their friends, that *Time*, by supporting the war in Vietnam, had irretrievably offended those young adults who were coming of age during the protest-filled days of the late 1960's and early 1970's. As a result, according to the theory, the supposedly more liberal *Newsweek* had enjoyed an audience windfall among this oldest group of Baby Boomers, one that persisted to this day. Without knowing it, of course, this editor was

hypothesising a cohort effect - a sort of permanent grudge against *Time* among people who were young adults during the height of the Vietnam era. Of course, if one accepted the notion that *Time* is more conservative than *Newsweek* (a dubious proposition, I think), then one might also expect *Time* to benefit in the long run from a well known age effect - that age effect that tends to make people more conservative as they grow older. Without cohort analysis, one could never evaluate these alternative hypotheses empirically. Happily, 20 years of Simmons data and the techniques that I will describe next make it possible to do so.

REWORKING DATA TABLES FOR COHORT ANALYSIS

The raw materials for this cohort analysis are presented in **Table 4**. This table juxtaposes the total audience penetrations of *Newsweek* and *Time* by age group, for every year that the Simmons survey is available between 1965 and 1983; as such, each datum in the table represents the fraction of the total adult population of the US that claimed readership of each respective news magazine in each of those years.

To get these data ready for cohort analysis, one must convert **Table 4** into . what demographers call a *standard cohort table* - that is, a symmetrical table in which the intervals between the points in time correspond to the intervals delineating the birth cohorts. Since most age groups in the Simmons data are grouped in 10-year intervals, the construction of a standard cohort table requires two modifications, both in the interests of symmetry. First, one must make the 18-24 year age group into a 15-24 year age group. In this instance, this bit of magic was accomplished simply by fiat, by assuming the same penetrations among 15 to 17 year olds as among 18-24 year olds.

Second, according to the conventions of cohort analysis, one should convert the annual penetrations estimates into 10-year estimates. Given that this would waste too much data and force us to try to fit a cohort curve to only three points in time (1965, 1975 and 1985), I developed a semi-symmetrical semi-standard cohort table, displayed here as **Table 5**. In this table, the estimates for each year are made by calculating the simple averages of penetrations of adjacent years. Thus the 1965 estimate is an average of 1965 and 1966, the 1970 estimate is an average of 1969, 1970 and 1971, and so forth. This procedure has the theoretical advantage of reducing the distorting effect of sampling variability in any single year, and the practical advantage of filling in years for which data are missing. Of course, one might want to test the robustness of the analysis by experimenting with different averaging algorithms; my own experimentation with this dataset convinced me that the adjacent year approach was quite satisfactory.

The final step in the conversion of raw data into cohort data involves 'pulling the cohorts' out of the standard (or, in this case, semi-standard) cohort table. **Table 6** accomplishes this step, but not without the application of one more smoothing and averaging technique that compensates for the slight asymmetry of a table with 10-year age intervals, but five-year time intervals If **Table 5** were fully symmetrical, one could pull the cohorts out simply by following along its diagonals and copying the necessary figures into the rows of **Table 6.** For example, those who were 15-24 in 1965 were 25-34 in 1975 and 35-44 in 1985; thus, following along **Table 5's** diagonal, we see that *Newsweek* had 9.8% penetration of this cohort in 1965 and 13.4% penetration in 1975, and 11.5% penetration in 1985. The problem of course, is that we have no estimates for this cohort for 1970 and 1980. To derive these, one can take simple average of penetrations in the adjacent age categories during the intervening years. Thus, the 1970 *Newsweek* estimate for this cohort is

TABLE 4
US penetrations of Time and Newsweek by age
1965 - 1983 (raw data)

	65	*66*	*68*	*69*	*70*	*71*	*72*	*73*	*74*	*77*	*79*	*80*	*81*	*82*	*83*
Newsweek															
Age															
18-24	9.3	10.2	10.9	12.7	15.1	15.5	15.3	15.3	18.0	14.3	13.8	12.5	10.6	11.5	11.7
25-34	8.1	9.6	9.7	10.7	11.3	11.4	12.9	13.3	16.0	15.3	14.2	13.5	11.7	11.4	11.0
35-44	9.3	10.1	9.1	9.4	10.3	11.4	10.6	10.2	14.0	11.7	10.7	10.5	12.0	10.8	11.7
45-54	8.0	8.6	8.4	8.9	9.6	10.3	10.0	9.3	14.4	12.0	10.2	11.0	9.1	11.5	9.0
55-64	6.5	6.8	7.4	8.2	8.8	9.0	9.3	8.2	10.2	9.6	9.4	7.3	7.7	8.6	9.0
65+	7.3	3.3	3.9	4.0	4.5	4.5	3.7	5.0	6.5	4.9	5.7	5.0	5.4	5.8	4.9
Time															
Age															
18-24	14.2	12.9	15.2	19.0	20.2	21.9	21.0	21.8	16.5	16.8	16.0	15.5	13.9	15.0	16.0
25-34	13.0	13.9	14.3	15.6	18.2	19.0	18.4	18.0	16.0	17.6	15.3	14.7	16.4	15.2	14.9
35-44	12.5	13.7	13.5	15.7	17.0	18.1	17.3	17.1	14.7	15.0	12.9	13.0	14.6	13.2	13.5
45-54	11.2	12.5	12.7	13.6	15.0	15.5	15.3	15.3	15.8	14.4	13.8	14.2	13.3	13.9	13.1
55-64	9.5	10.8	11.7	10.8	12.4	12.5	13.0	13.2	11.8	11.2	10.8	8.8	10.2	10.0	12.1
65+	6.8	5.8	7.6	6.7	8.1	6.0	7.0	6.2	6.7	6.7	6.7	7.6	7.3	7.7	7.1

the average of the 1970 penetrations for the 15-24 and 25-34 age categories. This approach makes the simplifying assumption that the cohort is distributed equally between the two age groups. While this assumption is not strictly true, experimentation with more complicated alternative procedures - including cohort component adjustments and logistic interpolation techniques - failed to alter the basic findings. Given these experimental results and my recognition that demographers are rare in the media business, I can recommend this simple adjacent averaging technique as a good approximation.

Once you have computed your own version of **Table 6**, you can then run to the computer graphic package and graph the results. Here is where the theoretical payoff comes. Returning to our editor's theory about the grudge against *Time* and the windfall for *Newsweek*, let us examine the cohort trajectories that can be graphed from **Table 6.**

TABLE 5
Time and Newsweek penetrations
semi-standard cohort table format

	1965	*1970*	*1975*	*1980*	*1985*
Age					
15-24					
Newsweek	9.8	12.9	16.0	13.5	11.3
Time	13.6	18.1	20.3	16.1	14.9
25-34					
Newsweek	8.9	10.6	13.4	14.3	11.4
Time	13.5	16.0	17.9	15.9	15.5
35-44					
Newsweek	9.7	9.6	11.6	11.0	11.5
Time	13.1	15.4	16.8	13.6	13.8
45-54					
Newsweek	8.3	9.0	11.0	11.1	9.9
Time	11.9	13.8	15.5	14.1	13.4
55-64					
Newsweek	6.7	8.1	9.2	8.8	8.4
Time	10.2	11.6	12.6	10.3	10.8
65+					
Newsweek	5.3	4.1	4.9	5.2	5.4
Time	6.3	7.5	6.5	7.0	7.4

What do we see here: age effects, cohort effects, or period effects? There are modest cohort effects: In **Figure 3** each successive cohort jumps on to the chart with a slightly higher *Time* penetration than the one before it, and it remains higher throughout the 20-year period. In fact, a demographer would say that the graph displays considerable 'cohort inertia' - just because the cohorts are differentiated enough that the cohort lines never touch one another. Age effects do not look too strong, though levels of penetration decline very very slightly with age for all cohorts (except, perhaps the youngest). What is surprising is the strength of the period effect that seems to be centred around 1974 - the year in which the Watergate crisis climaxed in the resignation of President Nixon. On the evidence of this chart, *Time* enjoyed a temporary spurt in its audience during the Watergate period, but has otherwise had rather gradual and normal success in winning audience among all four cohorts. There is no evidence here of a sharp rejection of *Time* among embittered Baby Boomers: indeed the top line (Baby Boomers) is higher than for any other cohort.

Figure 4 shows the cohort analysis results for *Newsweek*. Again, there are modest cohort effects and negligible age effects. Curiously, *Newsweek's* audience measures do not appear to have profited from the upsurge in interest in news around the time of Watergate. In terms of the *Time* editor's theory, there is nothing here to suggest a mass defection of angry Baby Boomers to *Newsweek* during the early 1970's. Indeed, *Newsweek* made its greatest audience gains among the Baby Boomers during late 1970's - well after the Vietnam war had faded from the picture.

The issue is perhaps best explored in **Figure 5** where the trajectories for the youngest two cohorts are plotted for both magazines. Here we see that both magazines have improved their penetrations with each successively younger cohort, both by comparable

TABLE 6
Time and Newsweek penetrations recast as cohort data

	1965	*1970*	*1975*	*1980*	*1985*
Cohort A (Born 1946-1955)					
Newsweek	x	12.9	14.7	14.3	11.5
Time	x	18.1	19.1	15.9	14.7
Cohort B (Born 1941-1950)					
Newsweek	9.8	11.8	13.4	12.7	11.5
Time	13.6	17.1	17.9	14.8	13.8
Cohort C (Born 1931-1940)					
Newsweek	8.9	10.1	11.6	11.1	9.9
Time	13.5	15.7	16.8	13.9	13.4
Cohort D (Born 1921-1930)					
Newsweek	9.7	9.3	11.0	10.0	8.4
Time	13.1	14.6	15.5	12.2	10.8

Key:
Cohort A = Ages 15-24 in 1970: b.1946-1955
Cohort B = Ages 15-24 in 1965: b.1941-1950
Cohort C = Ages 25-34 in 1965: b.1931-1940
Cohort D - Ages 35-44 in 1965: b.1921-1930

margins. If anyone had a windfall here, it was *Time*. *Time* widened its audience advantage considerably during the Watergate period; however this windfall was temporary, and by the late 1970's, the more familiar, historical pattern had reasserted itself. While one might want to puzzle over graphs like this to infer other conclusions or to speculate on other theories, I concluded from this graph that there was, at least, no *empirical* evidence to support the editorial hypothesis of a long-term grudge against *Time*.

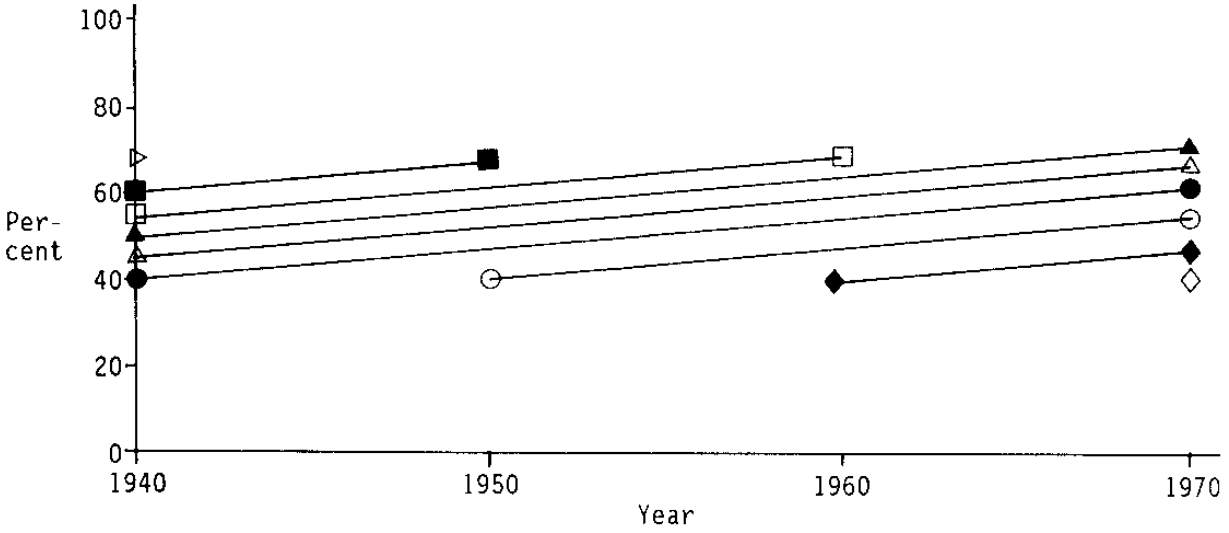
OBJECTIONS AND REFUTATIONS

Now, one may raise several very reasonable objections to the cohort analysis approach that I have mapped

out here. One may quarrel with the averaging methods that I have proposed, or with the assumptions that underlie them. One may worry that the data - riddled as they are with methodology changes, interruptions, inconsistent tabular category aggregations, universe redefinitions, and sampling variability - just cannot support such sweeping historical synthesis. Or one may simply think that cohort analysis is too arcane and academic to be of use to anyone.

While there is merit to these objections, cohort analysis still deserves our consideration. After all, the Simmons archive, imperfect though it may be, is, in the United States at least, our only really long-term, relatively consistent time series on magazine audiences. Thus it offers the only opportunity to resolve these larger historical and theoretical issues empirically. Considering the wealth of data in that archive, not only on narrow competitive issues, but also on trends in consumption of whole classes of products, there are wide possibilities for strategic analysis. Imperfect data are still better than no data, and empirical testing of hypotheses is still to be preferred over idle theoretical speculation. What I have sketched for you in this paper is a way to use an approach common among demographers to get a better picture of the age, period and cohort dynamics affecting long-term trends in media use.

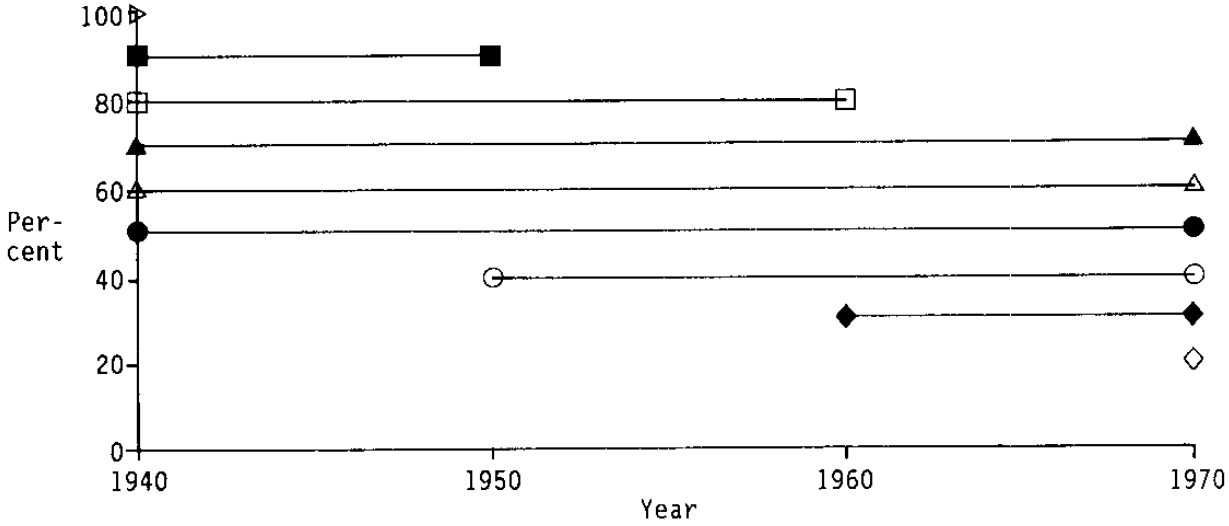
FIGURE 1
Chart of hypothetical cohort data in which all variation is due to age effects



Legend

◇ *Cohort A - b.1941-1950*
◆ *Cohort B - b.1931-1940*
○ *Cohort C - b.1921-1930*
● *Cohort D - b.1911-1920*
△ *Cohort E - b.1901-1910*
▲ *Cohort F - b.1891-1900*
□ *Cohort G - b.1881-1890*
■ *Cohort H - b.1871-1880*
▷ *Cohort I - b.1861-1870*

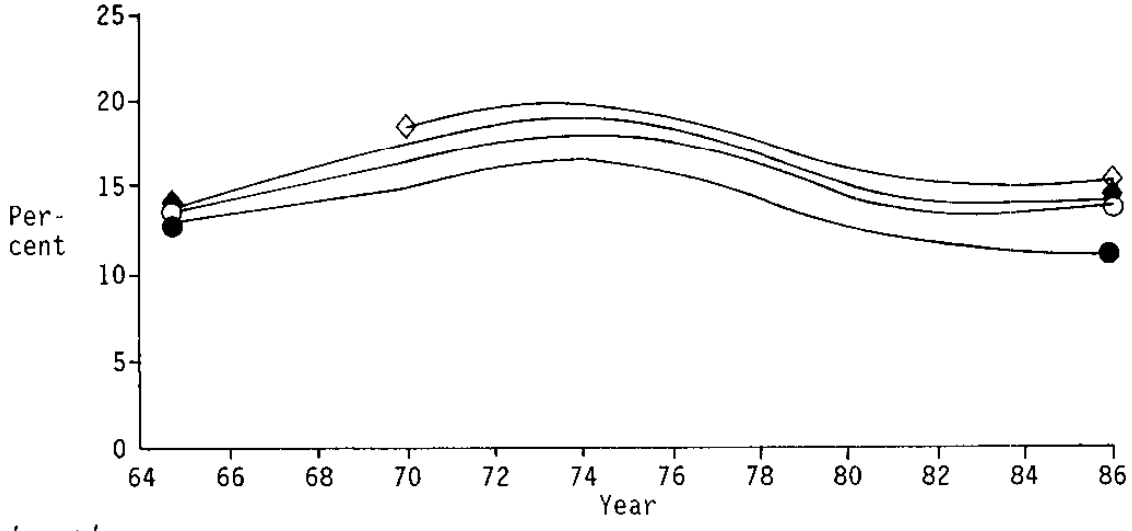
FIGURE 2
Chart of hypothetical cohort data in which all variation is due to cohort effects



Legend

◇ *Cohort A - b.1941-1950*
◆ *Cohort B - b.1931-1940*
○ *Cohort C - b.1921-1930*
● *Cohort D - b.1911-1920*
△ *Cohort E - b.1901-1910*
▲ *Cohort F - b.1891-1900*
□ *Cohort G - b.1881-1890*
■ *Cohort H - b.1871-1880*
▷ *Cohort I - b.1861-1870*

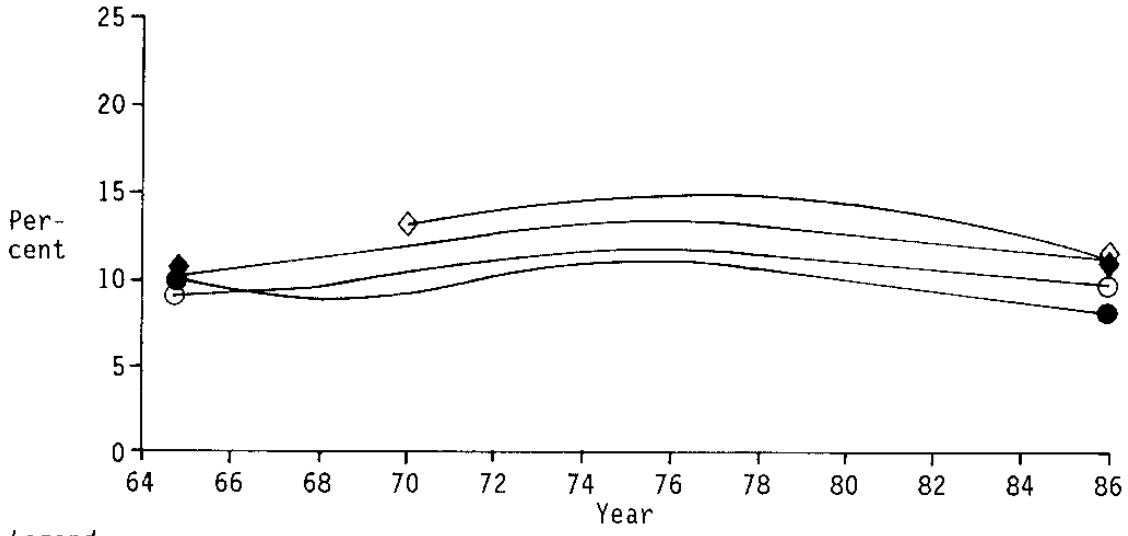
FIGURE 3
Cohort trajectories for Time adult audience 1965-1985



Legend

◇ *Cohort A - b.1946-1955*
◆ *Cohort B - b.1941-1950*
○ *Cohort C - b.1931-1940*
● *Cohort D - b.1921-1930*

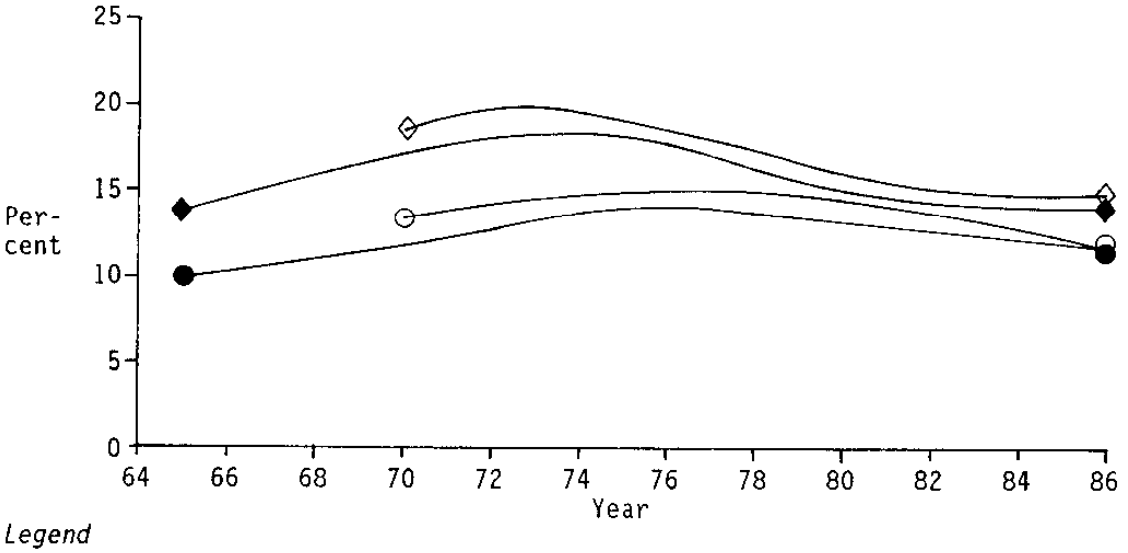
FIGURE 4
Cohort trajectories for Newsweek adult audience 1965-1985



Legend

◇ *Cohort A - b.1946-1955*
◆ *Cohort B - b.1941-1950*
○ *Cohort C - b.1931-1940*
● *Cohort D - b.1921-1930*

FIGURE 5
Cohort trajectories for adult audience
Time and Newsweek, 1965-1985



Legend

◇ *b.1946-1955 Time*
◆ *b.1941-1950 Time*
○ *b.1946-1955 Newsweek*
● *b.1941-1950 Newsweek*