



Birch Radio

BIRCH vs. ARBITRON **The Researcher's Perspective**

William E. Engel
Thomas F. Griffin, Ph.D.
Charles Palit, Ph.D.

November, 1988

Researchers Tackle Arbitron's
Behind the Numbers:
A Comparison of Arbitron and BIRCH

Arbitron's Agenda:

- I. Diary vs. Telephone**
- II. 7-Day vs. 1-Day**
- III. Pre-Alert vs. Cold Call**

The Researcher's Issues:

1. Sample Frame
2. Response Rate
3. Sample Size
4. Mode of Collection
5. Sample Design
6. Estimation Procedures

Reviewing Arbitron's Agenda:

I. Diary vs. Telephone

- Every collection instrument has its own strengths/weaknesses.
- BIRCH selected telephone because:
 - Highest response rate (50-60%)
 - Greater control over interview
 - Respondents do not have to read instructions
 - Interactive clarification
 - Probing
 - Uniformity of reporting across respondents

Reviewing Arbitron's Agenda:

Diary vs. Telephone (continued)

BIRCH rejected the diary because:

- Unacceptable response rates (30-40%)
- No control over interview
 - PULSE Study
- Literacy rate
- Requirement for multiple methodologies:
 - Sampling (Listeds vs. Unlisteds)
 - Incentives (age/sex/race)

Reviewing Arbitron's Agenda:

Diary vs. Telephone (continued)

A Word on Control:

- **PULSE Study - December 6, 1968**

Subject:

Unquestioning credence has been extended to the claim inherent by definition that the diary record of radio listening is concomitantly kept with the actual occurrence of the listening.

... the diary technique depends entirely upon this claim.

Test:

- Diaries placed by mail (Arbitron methodology followed)
- Respondents told to mail booklets at end of survey week
- PULSE retrieved diaries on fifth day (in person)

Results:

Blank days	42%
1-4 days Incomplete	11%
Filled out for all 7 days (after 4 days)	6%
Properly executed (filled out 4 days)	37%
Diaries mailed in (unable to pick up)	4%

Conclusion:

If Arbitron generates a 40% response rate ...
then "properly executed" response rate is 14.8%.

Reviewing Arbitron's Agenda:

Diary vs. Telephone (continued)

Examining Arbitron's Claims:

Arbitron Claim:

1. *People of all lifestyles fill out diary.*
2. *Almost all diaries are kept in a timely manner.*

Evidence Shows:

- Literacy rates and low Response Rates must negate this claim.

(A 30-40% response does NOT indicate a match of *ALL* lifestyle groups.)
- PULSE Study showed 37% filled out on correct day (that's not "almost all").
- Arbitron has never replicated PULSE Study and cannot back claim.

Reviewing Arbitron's Agenda:

Diary vs. Telephone (continued)

Examining Arbitron's Claims:

Arbitron Claim:

3. *Diary fits today's lifestyles.*
4. Mel Beville quote:
*... No higher quality survey technique
affordable by local radio has come
to being.*
5. *Arbitron compares its nationwide
results to RADAR.*

Evidence Shows:

- If claim is true, execution is awful, since Response Rates are 30-40%.
- Since Arbitron executes well, claim must be rejected.
- Quote from 1985, BEFORE BIRCH revised methodology in January, 1986, to compete.
- BIRCH was not shown.
- Arbitron changed a significant edit rule to drive levels UP to RADAR levels (after daypart diary introduction).

Respondents who list start times only are credited with 90 minutes of listening vs. 15 minutes (under old rules).

Ogilvy & Mather
findings:
June, 1987
- RADAR levels are trending down . . . diary, trending up.

Reviewing Arbitron's Agenda:

Diary vs. Telephone (continued)

We believe these problems are only part of the much larger issue of whether the diary itself, daypart or otherwise, provides the best means to estimate radio listening behavior. The fact that roughly 60% of those surveyed don't want to participate in a diary-based study, we believe, is a serious methodological shortcoming - and one which points to the telephone as a better means of gathering radio audience information.

Kevin Burns
John Hunt
Jon Swallen
Ogilvy & Mather June, 1987

Reviewing Arbitron's Agenda:

Summary: Diary vs. Telephone

- Telephone is not perfect, but:
 - It produces 50-60% response rates;
 - It offers control over the interview.
- Diary offers less:
 - Response rates 30-40%;
 - "Same-day" response rate closer to 15%;
 - Television has rejected diary.

Reviewing Arbitron's Agenda:

II. 7-Day vs. 1-Day:

Statement: *Seven-day measurement has positives and negatives:*

- In theory, it is optimal
- In practice:
 - It cannot be executed without direct interviewer contact.
 - Respondent fatigue major execution problem.
- BIRCH elected to measure a single day because:
 - Response rates;
 - High degree of day-to-day correlation (people listen similarly day-to-day);
 - Reporting fatigue is significantly reduced.
- BIRCH *DOES NOT* reject 7-day telephone measurement:
 - Cost is not practical for all markets;
 - Respondent fatigue is still a problem.

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

- Birch rejected 7-day diary measurement because:
 - Response rates
 - Practice doesn't follow theory:
 - Respondents do not fill out diaries every day (PULSE Study).
 - Reporting fatigue:

Arbitron Daily Cumes by Day of Week¹:

<u>Thursday</u>	<u>Friday</u>	<u>Monday</u>	<u>Tuesday</u>	<u>Wednesday</u>
88%	87%	85%	84%	80%

¹ Ogilvy & Mather, June, 1987

- BBM of Canada finds similar daily declines in levels with their diary technique.
 - Levels are highest on Monday, first day of the survey week.

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

A Word on Reporting Fatigue

- Every respondent to every research instrument has a fatigue factor:
 - We call it "termination point"
 - At this point, respondent either:
 - Stops participating totally (Refusal)
 - Stops participating mentally (Complete)
- The point:
 - Fatigue drives down response rates;
 - Fatigue can make data un-decisionable.

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

Examining Arbitron's Claims

Arbitron Claim

1. *Without a question, a 7-day technique is more reliable than a 1-day technique.*

Evidence Shows:

- No evidence presented to support claim.
(Could be correct for 15% Response Rate that seems to fill out diary each day.)
- Evidence shows: Fatigue bias in 7-day diary.

2. More Observations:

- Disregarding Response Rate and fatigue bias:

High intra-class correlation day-to-day causes negligible increase in effective sample size.

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

Day-To-Day Radio Listening Is Highly Correlated:

- People tend to listen to same stations;
- People tend to listen at same times;
- Effect of measuring multiple days:

**More observations of the same behavior;
More sampling points but not more reliability.**

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

The Statisticians Speak:

For several years it has been recognized that not all observations contain the same amount of statistical information, and that **a unit of response is not equal to a unit of information**. In most situations a sample of 3,000 observations in **100 geographic clusters of 30 observations each** will **contain less information** than a sample of **3,000 observations which is unclustered**, because units that are close to each other tend to be correlated. This situation has led to the concepts of clustering, the value of an observation, effective sample size, and design effect.

A similar type of clustering effect occurs in obtaining repeated responses or measurements from the same individual. In marketing research studies it is common to ascertain a person's behavior or consumption by asking what he did yesterday, the day before, and so forth, to a given number of days. The question arises as to **how much additional information is accrued by obtaining reports for several days in lieu of a single day**. Assuming there is no memory lapse, **the amount of incremental information added depends upon the intra-class correlation between days within individuals.**

For many **consumption and behavioral items that occur on a daily basis**, such as the number of ounces of soft drinks consumed, the number of cups of coffee drank, the amount of time spent watching television, or the number of pages read in the daily newspaper, **the intra-class correlations tend to be high**. If for a particular consumption the intra-class correlation is 0.7 and the value of a measurement on a single day is considered to be equal to 1.00, the amount of information contained in two days is 1.18; for three days it is 1.25, for four days 1.29, and for five days 1.32.

Martin R. Frankel, Lester R. Frankel, *Some Recent Developments in Sample Survey Design*, Journal of Marketing Research, #14 (August, 1977).

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

Examining Arbitron's Claims

Arbitron Claim:

3. *Arbitron claims that more sample means more reliability.*

4. *Arbitron shows a 500 lead in in-tab over BIRCH in San Francisco.*

Evidence Shows:

- This is correct.
- Their application of the claim is not correct.

Arbitron seems to believe that they can multiply their sample size by 7 measurement days.

Sample size counts each respondent only one time.

- This is correct
- San Francisco contains 2 other metros which are sampled as markets.
- Imbedded metros must be carefully considered in evaluations of sample size.

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

A Closer Look at San Francisco

- San Jose metro falls within the San Francisco metro geography.
- Santa Rosa also falls within the San Francisco metro.
- In-tab by area:

	Spring 1988		Fall 1988
	<u>ARB</u>	<u>BIRCH</u>	<u>BIRCH</u>
San Francisco	2,037	1,520	2,183
San Jose	1,430	1,578	2,190
Santa Rosa	<u>527</u>	<u>158</u>	<u>223</u>
Totals	3,994	3,256	4,596

- BIRCH has addressed in-tab differential by announcing a phased-in 70% increase of total in-tab for Top 10 markets by Spring 1989.
- Fall targets are listed above. Birch will be higher in Fall 1988.
- BIRCH has announced increases for weekends in markets 11-107 of 80% to 100% beginning in Fall 1988.

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

Imbedded Metros

- A caution when evaluating imbedded metro markets:
 - Total market sample size may not be used in standard error calculations:
 - Station spill over geographical areas;
 - Standard errors for estimates within oversampled counties will be different from those in non-oversampled counties.
 - *Example:* Station with 50% of listening within San Jose (Santa Clara County) and 50% within Alameda County:
 - That portion of audience within Santa Clara County must be evaluated separately from:
 - That portion of audience within Alameda.

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

Real Cumes

Question: *Since Arbitron represents the listening of a minority of the market's population (30-40% Response Rate) . . .*

How can the cume estimate be "real?"

- Birch elected to project to a model because:
 - High degree of intra-class correlation spelled high degree of reliability for a modeled approach.
(See "The BIRCH Weekly Cume: Radio's Best Estimate.")
 - Concern about fatigue, over time. A model will eliminate fatigue as a factor.
 - Measurement cost would be better transferred to greater sample size.
 - Model first explored by Audits and Surveys, who found their results "encouraging."
 - Model explored by BIRCH through Engel, Griffin and Yergin in separate development efforts.

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

Examining Arbitron's Claims (continued)

Arbitron Claim:

5. *Arbitron cumes are actual cumes based on 7 days of listening.*
6. *BIRCH cumes are modeled cumes, generated from a Statistical Model that was based on 7-day estimates.*
7. *Arbitron says "Fact is, modeled cumes don't work."*

Evidence Shows:

- PULSE Study reminds us that only 30% of the diary keepers maintain it up-to-date;
- The other 70% we assume to be recall and not actual.
- Birch cume model built on statistical logic, validated against 7-day estimates.
- Arbitron data run through the model correlates to in excess of .97 with their printed "real" cumes.

(Used Arbitron average daily cumes with average daily duplication to eliminate fatigue bias in result.)
- Modeled cumes for 1-day interviews match those produced by 7-day telephone.

(Birch, Madison, WI., Pre-Test, 1988)

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

Examining Arbitron's Claims
(Continued)

Arbitron Claim:

8. *It is impossible to have Saturday cumes larger than Monday-Saturday cumes.*

Evidence Shows:

- Previous to Spring 1988, BIRCH processed weekpart independent samples.
- While statistically correct, conflicts did exist.
- Clients requested that the data be conformed.
- Action: Problem no longer exists as of Spring 1988.

Note:

A similar situation exists within Arbitron's AID System. When demos are requested within county, projections often exceed the actual population for that demo within the county by 100% to 200%. At the same time, the same demo groups within other counties will be under-represented by the same amount.

The reason for the disparity: Arbitron allows the data to conform with that in the market report, forsaking accurate estimates in AID to the county level.

Arbitron's decision in 1976 to conform AID to the market report set the precedent for BIRCH to do the same with its weekends.

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day (continued)

Examining Arbitron's Claims (Continued)

Arbitron Claim:

9. *The BIRCH Teen Impossibility*

Evidence Shows:

- Arbitron says BIRCH over-represented Teen population by 25%.
- Problem: Arbitron used BIRCH's average daily duplication as a multiplier against the weekly cume. (Like adding two stations' weekly cumes to get a net reach!)
- Problem: Arbitron's oversight to this apples and oranges case.

Reviewing Arbitron's Agenda:

7-Day vs. 1-Day

Summary and Conclusions

The evidence says:

- That 7 days of data, if collected correctly, gain a negligible increase in reliability.
- That fatigue bias is difficult to overcome over 7 days.
- That response rate decline because of the 7-day measure is not worth the price.
- That money would be better spent on increasing sample sizes than approaching 7-day collection.
- That, because of the high degree of intra-class correlation that exists day-to-day within the individual, the perfect environment exists for a highly efficient and effectual cume model.
- That the BIRCH cume model's high degree of correlation with both diary and telephone 7-day measures makes it the optimal solution for estimating 7-day audiences.

Reviewing Arbitron's Agenda

III. Pre-Alert vs. Cold Call:

Statement: Pre-Alert is done to raise response rates.

It would be impossible to conduct a 7-day study without pre-alerting the sample.
(Sample must know, day-to-day, that you are going to call that day.)

Birch elected to use the 1-interview design without alert because:

- It is an accepted technique;
- Levels and estimates compare favorably with industry tests;
- Proper questionnaire design with interactive probes may actually out-perform pre-alert;
- "Halo Effect" around pre-alert has never been quantified for the medium;
- Improvement in Response Rates for single interview design.

(1986 Conklin Study of Pre-Alert on Radio Interview)

Reviewing Arbitron's Agenda

Pre-Alert vs. Cold Call

Examining Arbitron's Claims

Arbitron Claim:

1. *Enhances legitimacy of survey*
2. *Opportunity to focus on part of daily living that's usually taken for granted.*
3. *Helps insure that respondent will report only listening for the surveyed time period.*
4. *Courteous and business-like.*
5. *Cold-calling is cheaper but sacrifices quality.*

Evidence Shows:

- BIRCH's 60% response rates eliminate this claim.
- If this is their pre-alert method . . . it is called "biasing the sample."
- BIRCH estimates show that telephone probes work to uncover secondary and tertiary listening.
- Not what PULSE Study found!
- BIRCH's interviewers are also "courteous and businesslike."
- BIRCH has a 20-point + response rate over Arbitron.

Reviewing Arbitron's Agenda

Pre-Alert vs. Cold Call (continued)

Examining Arbitron's Claims

Arbitron Claim:

6. *Cold-calling creates problems - telescoping, and turning research into brand-awareness.*
7. *BIRCH shows top-rated stations with HIGHER estimates than the top-rated stations in Arbitron.*

Evidence Shows:

- Recall is a problem in the diary, too!
- Interactive benefit of the telephone is that probing takes place to aid recall.
- BIRCH interview screens out brand awareness AND voting.
- Read Arbitron Non-Response Study.
- Arbitron admits that it under-measures audiences to station formats that are most likely to have the largest 12 + shares.

Reviewing Arbitron's Agenda

Pre-Alert vs. Cold Call (continued)

Summary

- Seven-day methodologies MUST pre-alert in order to measure.
- "Halo Effect" around pre-alert has never been quantified for the medium. (AID runs show that it is a factor.)
- 30-40% Response Rates for technique do not support 7-day measurement under Diary Technique as a viable measurement alternative.
- Is everyone in the household cluster uniformly informed of their tasks before the survey measurement period starts?
- How are "kept" diaries instructed at the respondent level as far as an alert?
- Single Interview Technique is an accepted technique.
- Single Interview by phone allows interactive probing.

A Look at Response Rates

- Arbitron says:

Comparison of Response Rates can be misleading.

(Different rates are due to) different methodologies.

Amount of data collected

- Response Rates are the EQUALIZER when comparing the success or failure of differing methodologies:
- There is one standard for calculating response rates:

In-Tab

Eligible Sample Within Pre-Designated Sample

- The amount of data collected means nothing if that data is collected from a small (response rate) homogeneous group of active diary-keepers.
 - What about the other 60-70% of the population?
 - Does bigger in-tab make response rate bias go away?

A Look at Response Rates (continued)

- Arbitron states that Response Rates are going:
 - Up for Arbitron
 - Down for BIRCH
- Response Rates are going up for Arbitron, but . . .
 - Response Rates in the top markets remain in the 30% range.
 - This must mean that Response Rates in the smaller markets are going up.
- Response Rates have stabilized for BIRCH in the 57 to 59% range for all markets.
 - Response Rates for BIRCH in the Top 10 markets are significantly higher than Arbitron.
- Response Rates for BIRCH did decline for two quarters in 1987. Enhanced procedures implemented in Winter 1988 reversed this downward trend and produced some of the highest levels yet in the Top 10 markets.

Multiple Persons or Single Person Per Household

Responding to two Arbitron statements:

Multiple Persons Per Household is used in television because they generate household estimates, along with persons estimates.

Multiple Persons Per Household is used by the U.S. Census because they generate household estimates.

These two statements are apparently an attempt to equate "quality" with multiple persons measures, when in fact Arbitron Radio should be attempting to measure something other than household listening.

Arbitron states that MPPH is subject to "possible clustering." This should be modified to "unavoidable clustering." (See Frankel's article excerpt earlier.)

The BIRCH/Scarborough 1988 Research Agenda

1. Weighting Tests
 - Effects of weighting classes on variance and mean squared error
 - Effects of extended demographic weighting
 - Effects of modification of probability of selection weighting
 - Effects of weighting qualitative data
2. Replication Study
 - Calculation of Effective Sample Bases
 - Calculation of Statistical Efficiency
3. Study of No Answers/Unresolved Numbers
 - Effects on Response Rates
 - Validation
4. Seven-Day Collection of Data
 - Additional proof on Cume Model
 - Effect of using two-day interview on:
 - Cume Duplication (turnover)
 - Average Quarter Hour Stability
5. Response Rates:
 - Daytime Attempt Study
 - Multiple Attempts (5)
 - Rolling Sample Across Months
 - Modified Opening
6. Spanish Language Test
7. Black/Hispanic Oversample Test

The BIRCH/Scarborough 1988 Research Agenda (Continued)

8. Multi-Media Questionnaire
 - Radio/Newspaper Questionnaire
 - Product Booklet
 - Development of Cume Model
9. Questionnaire Design
 - Lifestyle Anchors
 - Occupation question
 - Localized Retail
10. Additional Coverage Validation
11. Factor/Cluster Analysis of Listening Variables
 - Are there distinct cluster groups that would replace format identifiers?
12. Ascription of Qualitative Data
 - Would it add stability and predictability to the estimates?
 - How to achieve?
13. Cume Model
 - Enhancement to daypart model
 - Fix weekend
 - Fix Cume Duplication section
 - Fix Exclusive Cume section
 - Employ formula rather than tables
14. Study of Performance in all measured markets
 - A. Levels
 - Two years worth of Quarterlies
 - What are trends? (vis-a-vis Arbitron)
 - Trends in qualitative categories
 - Trends by sex/age/demo
 - B. Spars trending
 - C. Age/sex distribution trending

The BIRCH/Scarborough 1988 Research Agenda (continued)

15. Joining Radio, Newspaper, Television, Product Usage:

- How do we conform data?

16. Study of effectiveness of each center:

- Spars comparisons
- Levels comparisons
- Ethnic comparisons
- Interviewer turnover
- Interviewer mix

17. Your ideas:

BIRCH/Scarborough Product Plans

1. Micro delivered database
2. Separate Newspaper Ratings Summary:
 - For the Top 75 ADIs
 - Use Existing Top 10
 - Survey September-May 1989 (yearly measurement?)
 - Shows single-source cross-media usage
 - Shows local retail shopping data by media vehicle
 - Shows product usage data by media vehicle
 - Single-source radio/newspaper/television ratings
3. Need modified methodology for Top 10 Multi-Media:
 - Review what we currently do
 - How to maintain highest quality with cost reductions (let's make money for a change!)
4. Increase in-tab in Top 10 markets by 70% for Fall
5. Continue modification of calling strategy in Top 10 markets to increase Response Rates
6. County Coverage

The BIRCH/Scarborough Agenda for Fall 1988 through Spring 1989

1. Research Projects
2. Weekend Sample Increases: **THIS FALL!**
 - Top 10 markets: 70% TOTAL in-tab increase by Spring 1989
35% in Phase I for Fall 1988
 - Markets 11-50: 80% increase in weekend in-tab, phased in beginning in Fall 1988
 - Markets 51-107 100% increase in weekend in-tab, phased in beginning in Fall 1988
3. Hispanic Procedures: High Density Hispanic Area (HDHA) implemented in Fall 1988 for Hispanic Report

Oversample placed in HDHA to meet SRAC guidelines.

Testing Spanish language weighting.
4. BIRCH/Scarborough Multi-Media Report:

Single source data for newspaper, radio, television, selected magazines.

Field work started in September for three markets, and October for 17 markets.

Standard BIRCH interview PLUS newspaper readership married to product usage/retail shopping booklet.

Reports due next June/July.

Markets 1-10 have just been released.
5. CountyAmerica (data for all counties in 48 contiguous United States) released for some states October 1st.

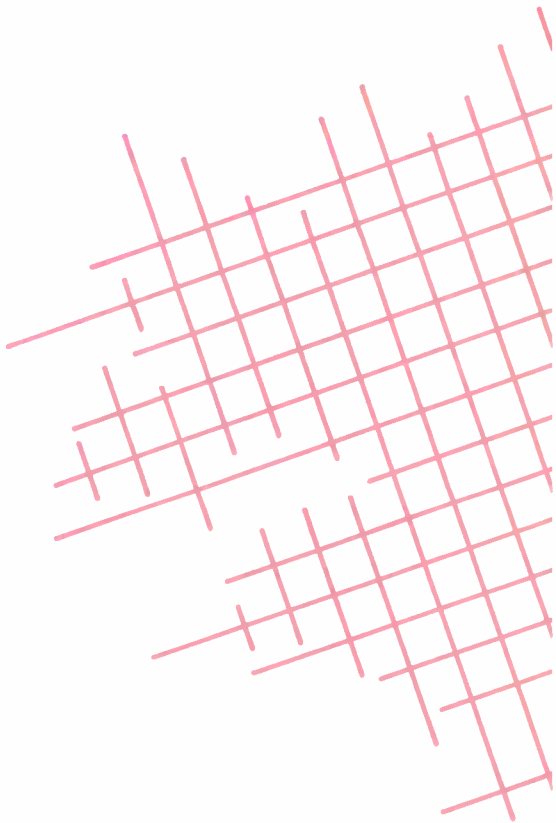
The BIRCH/Scarborough Agenda for Fall 1988 through Spring 1989 (continued)

6. Tulsa central interviewing facility on-line October 1st with 40 stations. Tulsa joins 5 other full-time BIRCH-owned central facilities.

Allentown, PA.
Coral Springs, FL.
Louisville, KY.
Sarasota, FL.
San Antonio, TX.

No work other than BIRCH/Scarborough interviewing is done out of these centers. Centers are staffed with BIRCH employees, no contract labor.

7. BIRCH interviewing up to 83,000 radio interviews per month.
8. EMRC running 6 months ahead of schedule.



Birch Research Corp.
12350 N.W. 39 Street
Coral Springs, Florida 33065
(305) 753-6043

Corporate Headquarters
44 Sylvan Avenue, #2D
Colonial Plaza
Englewood Cliffs, New Jersey 07632
(201) 585-7667

Regional Sales Offices

Atlanta
2110 Powers Ferry Road, #460
Atlanta, Georgia 30339
(404) 955-6317

Chicago
5105 Tollview Drive, #208
Rolling Meadows, Illinois 60008
(312) 392-5353

Dallas
14800 Quorum Drive, #465
Dallas, Texas 75240
(214) 934-3805

Los Angeles
18425 Burbank Boulevard, #415
Tarzana, California 91356
(818) 343-1244

Centralized WATS Interviewing Centers

Coral Springs, Florida
Sarasota, Florida
San Antonio, Texas
Louisville, Kentucky

Birch Radio is a research service of Birch/Scarborough Research