ARE WEB-BASED EXPOSURES CONTAMINATING PRINT MEDIA AUDIENCE ESTIMATES?

Scott C. McDonald, Time Warner Inc.

Two years ago at our meetings in Berlin, I warned that the rapid growth of the World Wide Web was introducing a potential new threat to the accuracy of audience measurement for print media. As publishers moved toward the creation of online versions of their flagship publications, consumers could in theory be exposed to electronic versions of publications that carry ads quite different from those found in the print versions. Moreover, given the global nature of the Web, there was, I argued, some potential for media exposure "spillage" across borders; this could result in further contamination of audience estimates for print media. Based on the limited data available at the time, I presented heuristic calculations of possible cross-border media exposures for a few European countries and concluded that the problem was not yet significant. However I urged that all audience research enterprises modify their source of copy questions to include the possibility of purely electronic exposures.

I return to this subject again today. In the intervening two years, a few things have changed. More publishers have created online versions of their print titles; initially these were "shovelware" that mirrored the print-version contact very closely; however now the electronic versions increasingly differentiate themselves from the print versions with unique copy and additional features. However the online editorial invariably carries the logo of the flagship print title and remains a source of potential confusion for respondents to readership surveys. Did I see it in Gourmet, or on Epicurious? Did I read it in People, or in People Online? If the advertisers supporting the print versions and the online versions were the same, it might not matter much. In fact, however, this is seldom the case.

On the measurement side, my somewhat informal canvass of practices in the field suggests that my advice in 1995 has been heeded only spottily. In the United States, MRI began last year asking respondents whether or not they ever read magazines online, and they have included online as a possible source in their source of copy question. To date, the reported levels of online reading of magazines has not been very high, according to MRI. More recently, SMRB added the option "Internet or online services" as a possible response category to its source of copy question in July of 1997; no data from that question are available yet. And in Hong Kong, the Hong Kong Media Index, conducted by AC Nielsen/SRG, has amended their definition of reading so as to explicitly ask respondents to **exclude** the Internet when thinking about their reading. Beyond these two initiatives, I am not aware of any other attempts by syndicated print audience research organizations to account for online sources as a potential source of copy or to jog respondents' memories about whether their claimed exposures were really print-based. I hope to stand corrected on this point by presentations at this symposium, but my impression thus far is that the issue is still not being taken too seriously by the syndicated audience research firms.

The Web itself has grown significantly in the United States and elsewhere. When we met in Berlin, estimates of the online population in the US -- that is the population accessing online services and/or the World Wide Web either from home, work or school in the past month -- ranged from 15 to 22 million. In 1995, there was a fairly distinct line separating the commercial online services (such as CompuServe and America Online) from the World Wide Web; by 1997 that line has all but vanished as the online services have integrated web browser functions. And with the release of the latest version of Microsoft's Windows '97, we shall see the web browser functions integrated into their ubiquitous desktop software.

The table below shows the projections for growth of online users from Jupiter Communications, a projection that represents a "middle range" estimate among the many available from various analysts.

Table 1
Projection of Total U.S. Online Users by Access Location, 1996-2002
(in millions)

	1996	1997	1998	1999	2000	2001	2002
Online Users Coming from Home	22.7	33.5	45.0	55.8	66.7	79.4	96.9
Incremental Business Users	9.1	8.4	9.0	8.4	6.7	4.0	2.9
Incremental Teens from School Only	--	0.1	0.3	0.8	1.4	2.3	3.0
College Students from University	5.0	7.2	8.9	11.0	12.6	13.0	13.5
Total Online Users	36.8	49.1	63.3	76.0	87.3	98.7	116.3

Source: Jupiter Communications, 1997

As you can see, the projections take account of the fact that most online users have access at both home and work: witness the relatively small (and diminishing) incremental importance of work-only users. Also notice that the projections assume increasing access from college students, a reflection of the growing ubiquity of computers and online access at universities and of demographic increases in the college-age population. Suffice to say that, from the "vehicle exposure" perspective, the potential for electronic contact with branded print media content will grow in the years ahead.

To some extent, these forecasts are conservative for they do not take into account the role of two relatively new factors on the digital media scene -- push technologies and corporate intranets. With push technologies, editorial content is automatically served to users without their specific request for that content. One well-known example of this at present is the PointCast screen saver which streams content and advertising to idle computers in screensaver mode. Publishers who cut deals with firms like PointCast can "push" their branded content to passive users. A host of new push software tools available for the first time this year give publishers greater flexibility to deliver their branded electronic content in novel forms to diverse clients. And many are now looking for opportunities to syndicate their content in the burgeoning Web marketplace. Beyond the Web, corporate intranets (or internal internets) represent another attractive syndication target: publishers get new paid outlets for their high-quality branded information and corporate CIOs are able to create more controlled, relevant information environments that ensure that online services enhance rather than diminish productivity.

My point is that developments in technology and in the online marketplace continue to expand the potential for confusion about the source of a respondent's exposure to branded editorial content.

How big of a problem is this really? One source of information on the extent of online exposure to print-based content comes from Media Metrix (or as it is more commonly known, PC Meter). PC Meter is a service that publishes monthly reach estimates for Websites, based on a randomly-selected panel of home-based Web users operating on Microsoft-based software systems. As such, it probably covers about 60% of the Web universe of today, most of which is in the Microsoft environment but some of which accesses the Web from places other than at home. PC Meter takes its measurements through a software program that queries what software or Website has the PC's focus of attention at any given second; in this regard it is a close cousin to the Nielsen people meter. It requires no recall on the part of panelists. The table below shows the PC Meter reach estimates for the Websites of several print-based publications for June of 1997. These are unduplicated reach figures; they refer to any visit to that Website by any individual in the panel households. If two household members visit the same site, it counts as two visits; if the same household member returns repeatedly to the site during the month, it still counts as only one visit. To make the reach estimates a little easier to comprehend, I have multiplied them by the Jupiter estimate of the number of 1997 online households.

Table 2
PC Meter Reach Estimates for Selected Websites
June 1997

	PC Meter Rating	Projected Unduplicated Audience Level
BUSINESSWEEK.COM	0.2	98,200
BYTE.COM	0.1	49,100
COVERGIRL.COM	0.1	49,100
EPICURIOUS.COM (Conde Nast)	1.0	491,000
FAMILY.COM	0.5	245,500
FOODWINE.COM	0.3	147,300
FORBES.COM	0.2	98,200
HOMEARTS.COM (Hearst)	1.5	736,500
NATIONALGEOGRAPHIC.COM	0.5	245,500
PATHFINDER.COM (Time Inc.)	7.1	3,486,100
PCWORLD.COM	2.2	1,080,200
PLAYBOY.COM	2.8	1,374,800
READERSDIGEST.COM	0.5	245,500
SAVVY.COM	0.1	49,100
SWOON.COM (Conde Nast)	0.2	98,200
TVGUIDE.COM	1.7	834,700
WIRED.COM	0.7	343,700
ZDNET.COM (Ziff Davis)	8.6	4,222,600

Source: PC Meter: World Wide Web Audience Ratings Report, Monthly Reach Trends, June 1997

As you can see, the audience levels are not enormous by the standards of print or television, but they are not negligible either. Having observed these ratings for some time now, I can tell you that they are fairly stable from month to month, with the general direction being upward. As these websites build their online audiences and as the online population continues to grow, the potential for confusion between online exposures and print-version exposures will probably grow.

Note that some of the noteworthy websites in Table 2 do not correspond directly to their print-based brands, but rather are online brands that host the online versions of that company's print titles. Examples include Conde Nast's EPICURIOUS (with content

from Bon Appetit, Conde Nast Traveler and Gourmet), Conde Nast's SWOON (with content from GQ, Details, Glamour and Mademoiselle), Hearst's HOMEARTS (with content from Redbook, Town and Country, Cosmopolitan, Country Living, Good Housekeeping, Popular Mechanics, American Home, Home Beautiful and Marie Claire), Ziff Davis's ZDNET (with content from Computer Life, Family PC, MacUser, MacWeek, PC Computing, PC Magazine, PC Week, and Interactive Week), and Time Warner's PATHFINDER (with content from Time, Life, People, Sports Illustrated, Sports Illustrated for Kids, Entertainment Weekly, Fortune, Money, Asia Week, Sunset, Southern Living, Cooking Light, Progressive Farmer, This Old House and Parenting). The fact that PC Meter reports only on the reach of the aggregated sites (i.e. the domain), rather than each component part, makes it difficult to use these data to evaluate the online reach of individual, print-based titles.

PC Meter also does not provide information on the extent to which individual exposures in any given month are correlated with exposures in previous months, nor does it provide any information regarding the correlation between online exposures and offline exposures from the flagship print publications. For insight into these dynamics, I turn to some internal data from Time Warner's Pathfinder on four titles -- a mix of weeklies and monthlies -- for a single week at midyear 1997. Data come from Pathfinder logfiles and from a series of online surveys conducted among users of different parts of the Pathfinder site.

Table 3
Weekly Visit Dynamics at Selected Areas of Pathfinder
June 1997

	Title A	Title B	Title C	Title D
Online page views per week	311,899	1,152,735	2,002,735	1,836,366
% of Total Pathfinder page views	2%	8%	15%	13%
Est # of unique visitors/month	69,722	278,888	522,915	453,193
% visiting site weekly or more (core)	47%	53%	58%	63%
% visiting site less often (light)	53%	47%	42%	37%
% subscribe to print version of title	25%	24%	8%	14%
% don't subscribe, but occasionally read the print version of title	50%	51%	65%	64%
% don't read print version at all, but read the online version only	25%	25%	27%	22%

Table 3 begins by showing the pageviews served by each of the online titles during a given week. Page views are the most fundamental unit of logfile data, and they are commonly reported by websites as an indication of relative levels of activity at different parts of their sites. If one looks solely at pageviews, the numbers look very large. However, the second row of Table 3 shows that, even with these impressively large counts of pageviews, the activity at each title's area still represented only a fraction of the total activity at the large and heterogeneous Pathfinder site. Nevertheless, the sheer weekly volume of pageviews supports the general argument that the growth of online represents a potential contamination of audience measurement for those print media that exist in both spheres.

But what is the relationship between pageviews and actual readers? Are there many users sampling a few pages each, or are there a smaller set of users ordering large volumes of pageviews. Alas, though I had hoped to be able to answer that question for you in this paper, I cannot -- at least with any exactitude.. To unduplicate visits across sessions and evaluate the volume of pages requested across sessions by individual users, websites need to use little digital devices called "cookies" -- little tags that are inserted onto a visitor's browser to enable the site to recognize them as a repeat customer the next time they come to the site. Users can refuse the cookies or can delete them in between visits, and some do so out of concerns about privacy. Pathfinder has only recently started experimenting with the use of cookies that persist for several weeks at a time, and it is not yet confident enough of the data's accuracy to release them. Instead, I present in the third row of Table 3 my own estimates of the unduplicated audience for each of the four titles, derived from applying their proportionate share of Pathfinder views to the PC Meter estimate for Pathfinder for June of 1997. Interestingly, these estimates are reasonably close to the preliminary estimates of unduplicated visitors obtained from Pathfinder's internal analyses.

In lieu of cookies-based data on repeat visits, the fourth and fifth rows of Table 3 present online survey data suggesting that, for each of the four titles, the online constituency is fairly loyal -- with between 47% and 63% claiming to visit the site weekly or more often. Pathfinder's own cookies-based analysis suggests that core loyal users request a disproportionate share of pageviews. However this analysis has not, as yet, taken account of the activity of intelligent agents and the "spiders" of search engines -- some of which have been shown to request as many as 500 pages in a single session. If we were to mistake these computerized agents for actual flesh-and-blood readers, their activities would mislead us into thinking that a hard core of users have absolutely nothing else in life to do other than request the pages of our websites. This is not likely. However it is likely that, once the cookies-based analyses are fully understood and digested, we will see that the estimated number of visitors to the title-specific parts of the Pathfinder site are far less than might be indicated by looking at the raw pageview numbers, but still not negligible. In other words they could still fuel some concern about confusion regarding the source of exposure to branded magazine content.

Another key question to consider in evaluating this issue is the question of whether online exposures occur instead of, or in addition to, print-based exposures. Again, the online survey data from Pathfinder can be helpful. Table 3 shows that, for two of the titles, roughly one quarter of the online visitors also subscribe to the print edition. This is testimony to the fact that the online editions of these titles provide new and different information -- more in-depth and timely information -- than their print-based versions; on the other hand, they do not offer the same splashy graphics and convenient portability of the printed versions. For these readers, the online versions complement the print-based versions. For the other two titles, the fraction of online users who also subscribe to the print version is significantly lower -- perhaps because these two titles are also more expensive as subscriptions.

Table 3 also shows that the modal group for all four of the titles is the group who do not subscribe, but who claim to read the print version occasionally. This group would include both newsstand buyers and pass-along readers. Presumably these readers switch back and forth between the print-based and online versions of these titles. They are the hardest group to study and, because of their size, potentially the most important source of possible confusion regarding the source of their exposure.

Finally, Table 3 shows that approximately one quarter of all visitors to these four areas of Pathfinder indicate that they never read the print version of these titles, but that they do read the online versions. This group could be termed a "pure incremental audience" for these titles. It would be interesting to observe their behavior over time to see whether their contacts with the online versions of the titles eventually spills over into incremental sales for the print-based versions.

In Table 4, I have provided an estimate of what, if the survey data and the PC Meter data are accurate, we might find as the actual level of incremental exposures to these four titles in one month. The estimate of "pure incremental audience" is merely the product of the fraction of survey respondents who say that they do not read the print versions but that they do read the online versions, multiplied by the PC Meter-derived estimate of the unique audience for that title for that month. It is more difficult to estimate the incremental audience contribution of the larger group who do not subscribe but occasionally read the print versions of the publications. For purposes of this illustration, I have arbitrarily assumed that 75% of this group really does read the print version of the titles at least one out of 4 times, and that they would legitimately be counted in the audience of the print versions of the titles. Accordingly, the estimate of total incremental audience counts only 25% of this group together with the "pure incremental" group.

Table 4
Illustration: Estimates of Incremental Audience
June 1997

	Title A	Title B	Title C	Title D
Est # pure incremental readers/month	17,431	69,722	141,187	99,702
Est # total incremental readers/month	26,146	104,583	206,551	156,352

Obviously, these estimates are not to be taken literally, but merely as an illustration of what magnitudes we might be dealing with at this point in time. I'll leave it to others to judge, but I would say that this is not a critical problem at the moment. However it should not go ignored.

Online media are growing and, despite plenty of fits and starts, they are unlikely to go away anytime soon. I see no evidence that they will displace the existing print media, but they may make the messy business of measuring print audiences even messier. Readers in the future will be presented with an increasing diversity of branded editorial content from many sources delivered to many destinations -- printed in magazines, retrieved over the Web, pulled by their personal agents, pushed by publishers to their TV sets or to their work-based intranet homepage or to their PCS device. At some point, their already taxed memories about where they saw that logo or that article may become unduly strained. At a minimum, we need to take steps in the source of copy questions to try to jog those memories as much as possible.

In the longer run, we probably cannot outpace the threat of location confusion. Whether this argues for abandoning methods which rely upon recalled behavior in favor of more integrated behavioral measures -- such as those that have been proposed in the past by Steve Douglas or such as are used by Nielsen and by PC Meter -- I cannot say. Perhaps it argues in favor of fusion techniques and such circulation-based measures as have been advocated recently by Paul Donato. On these matters I remain undecided.

What I am convinced of is that what I am calling location confusion should be added to the roster of measurement problems to be studied and considered by this body and by the syndicated research firms represented here. So in case anyone has grown weary of worrying about response rates, respondent burden, read to screen ratios, telescoping, and other well-established problems in readership research, I offer a new concern. If I am at all right, we shall be hearing more, not less, about it in the future.