The Annenberg Coding of Health and Media Project: Rationale and Plans

Bruce Hardy, Patrick E. Jamieson, Dan Romer, and Kathleen Hall Jamieson

Annenberg Public Policy Center

University of Pennsylvania

December 2008
The use of electronic media is a daily ritual for most Americans. Music, movies, television, internet, DVDs, VHS videos, radio, and video games are ubiquitous elements in today’s American household. For example, on a typical day, 83 percent of children under the age of six use some sort of screen media. The average time children of this age spend with screen media is roughly an hour and a half each day (Rideout & Hamel, 2006).

Consequently, the examination of the influential role that mediated content plays in shaping thoughts, directing affect, and prompting action has become the focus of an entire research discipline. One of the more robust findings from this research is that mediated communication is influential because it models behaviors. Humans have an “extraordinary capacity for symbolization” that allows them to “gain understanding and expand their knowledge by operating symbolically on the wealth of information derived from personal and vicarious experiences” (Bandura, 2001, p. 267). Many of these vicarious experiences come from exposure to mass media.

In order to understand the impact of electronic media on individuals, researchers must conduct a systematic analysis of its content. Despite their important contributions to our understanding of media, past analyses have been limited in scope. In particular, communication scholars have focused research resources on single behaviors modeled in a single medium over short periods of time. For example, the National Television Violence Study (1998) examined only three years of violence on television. In addition, most projects that code media content for violence do not examine substance abuse or other risky behaviors that may either lead to or follow from violence. The result is a
fragmented literature that fails to capture the complexity of risky behaviors and trends in these behaviors across media over time.

To address these concerns, the Annenberg Public Policy Center at the University of Pennsylvania in collaboration with the Robert Wood Johnson Foundation has initiated an integrated content analysis research program, the Annenberg Coding of Health and Media Project (ACHAMP). This project will content code both risky and pro-social/protective health behaviors modeled in popular film, television, music videos, radio, and the internet. A review (discussed below) of past content analyses of risky and protective health behaviors portrayed in media demonstrates the need for this project. In particular, the project will attempt to develop:

1) an integrated approach to content coding a wide range of risky and protective behaviors;

2) coding instruments that have passed high standards of intercoder reliability applied to the same coding units to permit broad use and ready replication and extension;

3) analyses of past media content that capture long-term trends;

4) analyses conducted across media to permit comparisons by media type; and

5) an online facility that allows public access to research reports and coding data for use by researchers, advocates, and policymakers.

Before we outline these goals in greater detail, we provide a brief overview of content analysis as a research methodology.
Content Analysis

Although scholars recognize differing definitions of media content analysis, for the purposes of this paper we will say that “Content analysis may be briefly defined as the systematic, objective, quantitative analysis of message characteristics” (Neuendorf, 2001, p. 1). Content analysis has been a main tool for scientific communication inquiry since the early propaganda studies conducted during the first half of the 20th century (e.g., Lasswell, 1927). Although there are a variety of research questions that can be answered using this methodology, for this paper we will focus specifically on content analyses that have been used to examine media effects.¹

The basic procedural framework for this method is well established. First, as directed by the research question, the investigator selects a population of media content from which to draw a sample. For example, this could be top-grossing films from 1950 to present, top-40 radio singles in 2006, all movies that feature teens in leading roles, and so on. Next the researcher identifies the unit of analysis. According to Krippendorff (2004a), there are three distinct units in content analysis: the sampling unit, the recording/coding units, and context units. “Sampling units are units that are distinguished for selective inclusion in an analysis” (Krippendorff, 2004a, p. 98). This is the unit of analysis that is used to generate a sample, e.g., a single film. The recording/coding unit is a specific section of the sample unit that will be analyzed. For example, in a study examining the portrayal of smoking in films, the sample unit is the film and the recording/coding unit is a five-minute segment within the film in which the act of smoking is depicted. The context unit is the unit of analysis “that sets the limit on the information to be considered

¹ Holsti (1969) suggested a three-part typology of the different research endeavors that one can accomplish with content analysis. Researchers can use this method to assess the meaning of messages, to look at antecedents of messages, and, finally, to examine the influence of messages.
in the description of recording units” (Krippendorff, 2004a, p. 101). Using the same example, the context unit would be the entire scene in which the smoking incident occurred.

The third and perhaps most important step of content analysis is the development of a coding scheme that is used by trained coders to reliably and validly classify messages within the sampled units (see Krippendorff, 2004). Most studies of media effects categorize gross characteristics of messages as well as the overall frequency at which such content occurs. For example, categorizations often used in studies of smoking in the media include: the frequency at which smoking is depicted, the characters shown smoking (further categorized by socio-demographic variables, the prominence of the characters, and so on), and the depicted consequences of smoking.

All content analyses should carefully apply measurement theory. Neuendorf (2001) describes the need for content analyses to have the appropriate reliability, validity, accuracy, and precision. She also emphasizes the importance of external validity, face validity, criterion validity, content validity, and construct validity.

**Intercoder Reliability**

A main goal of content analysis is to evaluate characteristics of messages as reliably as possible. “Intercoder reliability is the widely used term for the extent to which independent coders evaluate a characteristic of a message or artifact and reach the same conclusion” (Lombard, Snyder-Duch, & Bracken, 2002, p. 589). As Krippendorff (2004b) put it, “agreement is what we measure; reliability is what we wish to infer” (p. 414). The fundamental question is: Can we rely on the data obtained from content analyses to make generalizations and draw conclusions?
The standard by which one computes intercoder reliability in content analysis is contested (see the exchange between Lombard, et al., 2002, 2004, and Krippendorff, 2004b). Content analyses vary in their procedures for assessing intercoder reliability because there are many coefficients to choose from: simple agreement or percent agreement, Cohen’s (1960) $\kappa$ (kappa), Scott’s (1955) $\pi$ (pi), and Krippendorff’s (2004a) $\alpha$ (alpha). Each has unique properties. As we argue below, the wide range of approaches has led to some uncertainty about the validity of the findings that have been reported.

In the following pages, we provide an overview of past studies that have conducted content analyses of health behaviors portrayed in media. There is special treatment of the methodology of these studies as well as attention to gaps in the literature. Based on this review, we argue that an integrated approach to content coding of risky and protective behaviors is needed. This program of research should use coding instruments that meet high standards of reliability that can be applied across a wide time-frame and across different media so that trends can be captured. Finally, research reports and the data they are based on need to be widely available online to researchers and other interested parties. This is not intended to be a comprehensive review of all studies, but an illustrative overview that provides justification for our calls.

The Need for an Integrated Approach to Content Coding of Risky and Protective Behaviors

Most studies that have content coded mediated messages focus on one behavior in one medium. Coding of multiple behaviors is not the norm. The reasons vary. Research organizations and funding agencies tend to focus on a single behavior and researchers often receive funding support for specific research questions. For example, the work by
Stanton Glantz (e.g. Charlesworth & Glantz, 2005; Glantz, Kackirk, & McCulloch, 2004) and his colleagues focuses on smoking. Glantz, a University of California, San Francisco (UCSF) professor of medicine, is also director of UCSF's Center for Tobacco Control Research and Education and initiated the Smoke Free Movies project. His research is funded by the National Cancer Institute.

Ours is not an indictment of researchers who want to answer specific research questions. The results of these research endeavors are valuable. Instead, our project is designed to facilitate integration. Even the most thorough and comprehensive studies are limited in scope. For example, two of the more comprehensive content analyses of media, the Cultural Indicators Project and the National Television Violence Study, were limited to violence on television. Although these studies were massive undertakings, they examined only a narrow range of behavior and media. We believe that a comprehensive understanding of the influence of mediated portrayals on young people will not be forthcoming unless we examine the full range of modeled behaviors as they co-occur. For example, do instances of violence or suicide co-occur with abuse of substances? Do uses of substances themselves co-occur, and have these patterns changed over time? Do protective actions, such as discouragement of substance use or other risky behavior, co-occur? These questions illustrate the need for a more integrated examination of the co-occurrence of risk behaviors that we know are exhibited in the behavior of young people (Romer, 2003).

of films (200 movies from 1977-1988), content coded a variety of behaviors – cigarette smoking, over-the-counter or prescription drug use, illegal drug use, alcohol consumption, leisure-time physical activity, and poor eating habits – making theirs one of the more comprehensive studies looking at health behaviors in media. Durant and his colleagues (1997) conducted a content analysis of the portrayal of smoking and drinking in 518 music videos aired in the late spring of 1994 on the cable stations MTV (Music Television), VH1 (Video Hits 1), CMT (Country Music Television), and BET (Black Entertainment Television). A review of films by Trocki and Thompson (1993) connected alcohol use and sexual encounters. However, these authors did not conduct a systematic content analysis; instead, they provided their own interpretation of films that they had viewed.

Although a handful of studies have looked at multiple behaviors, there are still many gaps in the literature. Connecting and comparing behaviors in media is difficult because we must piece together past studies in order to figure out what we know about the portrayal of health behavior in media. Even if we are looking at a single behavior there is no ready way to compare the portrayal of that behavior across media by relying on the existing literature. For example, content analyses on smoking in film (e.g., Charlesworth & Glantz, 2005; Dalton, et al., 1997; Everett, Schnuth, & Tribble, 2004; Glantz, Kacirk, & McCulloch, 2004; McIntosh, Bazzini, Smith, & Wayne, 1998; Mekemson, et al., 2004; Omidvari, et al., 2005; Sargent et al., 2001; 2002; Stockwell & Glantz, 1997; 1998) greatly outnumber content analyses on smoking in any other medium, even television. Therefore, comparing the portrayal of smoking in television and movies by reviewing past literature becomes a difficult task, and may be impossible if
one is interested in other forms of media such as radio or the internet (See Ribisl, et al., 2003, for a content analysis of smoking culture and lifestyle websites listed on Yahoo!) – two media in which there is a dearth of content analyses of any kind, not to mention health behaviors.

**The Need for Reliable Coding Instruments Applied Consistently to the Same Units of Analysis to Enable Comparisons Across Time and Media**

If syntheses of coding studies found in the literature were attempted, one would find that sampling differs widely across studies, units of analysis differ, and coding instruments are rarely comparable enough to permit inferences. Furthermore, the reliability of the codes employed often may be subject to question. Many studies do not provide details of the coding instrument, making the task even more difficult. The definitions of categories used in the analyses often lack information required to replicate the coding. For example, Martino and colleagues (2006) conducted an analysis of sexual content in music lyrics. These authors provide little information on their coding procedure or their coding instrument:

Two raters independently coded the lyrics, obtained from Internet web sites, of all songs \( N = 193 \) from each of the 16 albums [from the 16 artists in their sample]. The unit of analysis was the song. Raters first judged whether a song contained 1 or more references to sexual behavior (implicit or explicit references to intercourse, oral sex, or other sexual acts). For each song deemed to contain \( \geq 1 \) sexual reference, raters then judged whether the song only contained \( \geq 1 \) degrading sexual references. Thus these classifications of content were mutually exclusive, and the degrading/nondegrading designation accounted for all of the instances of sexual content (Martino, et al., 2006, p. 434).
The best definition of what was considered a degrading versus nondegrading sexual reference was provided by way of two examples: lyrics from the band 98 Degrees and from the rapper Ja Rule.

Comparing the results of this study with another, for example Pardun, et al.’s (2005) study of sexual references in different media, would be difficult due to the fact that we could not determine whether their measures were comparable. This would be true even though Pardun and colleagues provided a coding sheet and a detailed description of their coding procedures. As political scientist Gary King wrote, “The only way to understand and evaluate an empirical analysis fully is to know the exact process by which the data were generated and the analyses produced” (1995). Yet, the full coding procedures are rarely detailed in research reports (e.g., Charlesworth & Glantz, 2005; Hazan, Lipton, & Glantz, 1994; Martino, et al., 2006).

Studies vary widely in their selection of units of analysis. For example, Stern (2005) used an entire film as the coding unit in her study on substance use in teen-centered films, while Terre, Drabman, and Speer (1991) used five-minute segments of the movies in their sample as the coding unit in their study on health-related behaviors in media. Dalton and her colleagues (2002), in their examination of smoking in films, used the entire film as the coding unit in a first step to identify occurrences of smoking, then “each occurrence of tobacco use or handling was then viewed multiple times to characterize the uses and its context” (p. 517). Stockwell and Glantz (1997) used five-minute segments to examine smoking in popular films. Different units of analysis will, of course, have different results.
These problems make it increasingly difficult to argue that new work shows either continuity with or a change from previous work. It is therefore difficult to correctly identify differences between newer and older studies as change in the portrayal of a particular behavior or an artifact of non-standardized codes. Changes in media over time cannot be assessed if variance exists among coding schemes.

A review by Kunkel, Cope, and Biely (1999) that compares three studies of sexual messages on television provides one of the few examples in which different studies “share the same conceptualization and operational definition of sexual content” (p. 231). As these researchers note, such a comparison provides a “unique opportunity to identify clear patterns” (p. 231).

Even when the content of coding systems can be compared, investigators often use weak measures of intercoder reliability. For example, simple agreement indices, uncorrected for chance, should not be used to assess reliability. Krippendorff (2004b) notes this problem:

The scales of chance-corrected agreement coefficients are anchored at two points of meaningful reliability interpretations, 0 and 1; whereas percent-like agreement indices are anchored at only one, 100%, which renders all deviations from 100% uninterpretable as far as data reliability is concerned…I suggest that the convenience of calculating percent agreement, which is often cited as its advantage, cannot compensate for its meaninglessness (p. 413).

Nevertheless, percent agreement indices are often used to assess reliability in studies. For example, Dalton and colleagues (2002) content analyzed the 25 top-grossing films from 1988-1997 for incidence and context of smoking. Their assessment of reliability, based on a simple agreement calculation, is questionable, thus rendering their conclusions
equally questionable. The researchers stated that “Two coders were selected and trained to conduct the content analyses. To evaluate interrater reliability, 10% of the movies were coded by both coders. All measures reported in this paper had a minimum percent-agreement of 70%” (p. 518). The first problem here is that they did not report the reliability for each category. On a positive note, the authors reported the minimum percent agreement in their study. Some studies report an average or an overall reliability, which may hide unreliable categories. Another potential problem with the Dalton, et al. (2002) study is that they did not use a chance-corrected reliability coefficient. A percent agreement of 70 could translate into a much lower coefficient when corrected for chance agreement to the point that some of the categories could be deemed unreliable. As a result, we do not know if the conclusions drawn by these researchers are capturing the phenomenon of interest or reflecting mere chance that coders agreed on some categories.

Some researchers do not even assess the reliability of the data. Consider Stockwell and Glantz (1997): “All films were coded by a single person [Theresa F. Stockwell]; we used this approach because our earlier study [Hazan, Lipton, & Glantz, 1994] showed good agreement between two independent reviewers” (p. 282). This is problematic for a variety of reasons. First, the person who coded the films is also the lead author of the article. Coders should be unaware of the hypotheses that motivate a study to reduce any possible biases that may influence coding decisions (Krippendorff, 2004a). Second, the study that these researchers rely on reports a “per film reliability ratio of .92 (SD = .07)” (Hazan, Lipton, & Glantz, 1994, p. 998), a percent agreement index.

A final concern in comparing coding systems across studies is variation in the underlying theoretical scheme used to define the codes. Different coding instruments
produce different results. This problem, which goes beyond standardization and reliability of coding instruments, is rooted in differences in theoretically driven conceptualization and operationalization of what is being coded. The coding of violence is perhaps the most difficult example of this problem.

Like many constructs in the social sciences, the meaning of violence is socially constructed, resulting in a variety of operational definitions found across studies. Depending on how one defines violence, a typical hour of prime-time television either contains 5.4 instances or as many as 38 (Potter et al., 1998). This illustrates the influence of operationalization on the reported results.

Not only do conceptions of what constitutes violence differ, individual responses to mediated violence differ as well (Dorr & Kovaric, 1980; Berkowitz, 1984; Tamborini, 1991). The judgment of whether “injurious behavior will be perceived as aggressive or otherwise depends heavily” on interpretations “of intentions and causality. . . .The same harmful act is perceived differently depending on the sex, age, attractiveness, status, socio-economic level, and ethnic background of the performer. As a general rule, people judge the harmful acts of favored individuals and groups as unintended and prompted by situational circumstances, but perceive the harmful acts of the disfavored as intentional and personally initiated” (Bandura, 1978, p. 12). Children’s ability to understand mediated violence also differs by age (Cantor & Sparks, 1984).

How violence is interpreted can have a significant impact on the coding procedure and, thus, working from a common definition is crucial. This becomes even more critical when we consider that different types of violence produce different effects.
More aggressive behavior is elicited when the violence that is observed is perceived to be real, not fictional (Van der Voort, 1986; Berkowitz & Alioto, 1973).

In devising a coding system, a researcher must rely on both past content analyses and effects literature to produce theoretically driven conceptualizations and operationalizations of the constructs that are to be coded. There is an interplay between content analytic studies and effects research that is sometimes ignored in the development of coding instruments. In developing codes for an integrated analysis, we must be aware that codes that have high reliability may still have different meanings for different audience members.

**Media Content Needs to Be Analyzed Over Time So That Trends Can Be Captured**

In 1970 communication scholar George Gerbner wrote, “We know next to nothing about trends in the composition and structure of mass-produced message systems that govern men’s lives and inform men’s minds in urban societies” (Gerbner, p. 70). Over 35 years later we still do not know much about trends in media content. This is in part due to the non-standardization of codes that we have just outlined; more importantly, there is simply a lack of comprehensive trend data available.

A handful of studies have collected some trend data on specific behaviors (e.g., Glantz, Kacirk, & McCulloch, 2004, on smoking). Possibly the most complete content analyses assessing trends in media were conducted by Gerbner’s Cultural Indicators research team – Gerbner, Larry Gross, Michael Morgan, and Nancy Signorielli. These researchers tracked the frequency of violence in prime-time television from 1967-2002. They found that violence in television remained quite high throughout those 30 years (see Gerbner, Morgan, & Signorielli, 1994; Signorielli, 2003). However, the Cultural
Indicators Project is the exception and not the norm. Most studies rely on much shorter time spans. For example, the National Television Violence Study consisted of three yearly samples (1994-1995, 1995-1996, and 1996-1997). Although the sampling frame and breadth of television shows analyzed were impressive, short-term coding cannot capture long-term trends.

Overall, there is a lack of consistent trend data on health behaviors portrayed in the media. Content analyses of sex in media are illustrative. Communication scholar Bradley Greenberg, for example, examined sexual activity in soap operas and compared findings from a 1985 study (Greenberg & Alessio) to a content analysis of soap operas from 1994 (Greenberg & Busselle, 1996). Comparing such results cannot produce conclusions about shifts in media, something these researchers acknowledge: “The most superficial conclusion from this study is that soap sex increased across the decade examined” (Greenberg & Busselle, 1996, p. 160).

Communication scholar Dale Kunkel and his associates (Kunkel, Eyal, Finnerty, et al., 2005a) conducted an impressive study of sex on television that collected data in 1998, 2002, and 2005. These scholars also recognized the lack of trend data on the portrayal of sex on television. They stated, “No previous program of research on sexual content has attempted to track such changes by applying the identical measures to samples of programming gathered across multiple points in time” (p. 57). Consequently, these scholars are in a better position to draw conclusions on shifts in media than the other studies outlined above. However, there is still a five-year gap between the first and second data collection periods and a three-year gap between the second and third data collection periods. There is the possibility that their conclusion that sex on television has
been steadily increasing is misleading. Figure 1 outlines the overall percentage of television shows with sexual content found in the Kunkel, et al. (2005) study and a fictitious, but possible, “true” trend for shows with sexual content.

As Figure 1 illustrates, the lack of complete trend data may hide “true” patterns. Our example illustrates the possibility that conclusions drawn from data for which there were large time spans between data collection periods might be misleading, or at least not telling the whole story.

Figure 1. Comparison of hypothetical and observed trends in sexual content on television.
Referring back to Gerbner’s call for a set of cultural indicators that tracks changes in culture, and not just violence, collecting and integrating trend data on changes in media content over the past half-century would open up avenues for scientific inquiry that may currently be closed. Conceptualizing such trend data as an indicator much like an economic and social indicator is useful in untangling the causal mechanism between mass media and society---i.e., does society mimic the media or do media mimic society? Using such trend data in conjunction with already established economic and social indicators is what George Gerbner envisioned over 35 years ago.

**We Need to Code Across Media and Different Types of Media**

Most content analyses on the portrayal of health behaviors in media have focused on film and television. Some studies have looked at a variety of media (e.g., Pardun et al., 2005). However, these studies are few and far between. We argue that to assess cultural content and its influence on behavior we need to code widely viewed music videos, music, and digital content, as well as television and movies. For instance, the internet is growing as the media of choice for Americans. In fact, currently 73 percent of Americans have access to the internet (Madden, 2006). Yet there are few content analyses of health behaviors on the internet. Other media, such as music and music videos, have only received attention from a handful of researchers (e.g., Baxter et al., 1985; DuRant et al., 2003; Pardun et al., 2005; Martino, et al., 2006; Smith & Boyson, 2002).

Consider content analyses of smoking in media. Most of the literature on the media portrayal of tobacco use has focused on film. Few studies (e.g., Breed & DeFoe, 1984; Christenson, et al., 2000; Cruz and Wallack, 1986) have examined the portrayal of tobacco use in television even though widely successful shows such as *Sex and the City*
and *The X-Files* featured main characters who smoke (see Borio, 2004). *Sex and the City*’s lead character Carrie Bradshaw smoked through most of the show’s run; in *The X-Files* the arch villain was dubbed “cigarette smoking man” or sometimes called “cancer man.”

Part of the problem with conducting content analyses of these other media is that before doing so, researchers must overcome many methodological obstacles. However, these other media hold the potential for research opportunities that might have been non-existent a few years ago. Consider the internet. It is a massive, decentralized network of hyperlinked multimedia contained in over 4 billion webpages on 400 million hosts (Internet Domain Survey, 2004). The nature of the internet presents unique challenges for communication scholars employing systematic content analyses to answer their research questions about content found online. Designing a sampling frame out of billions of pages is a monumental task. On the other hand, there is gigabyte upon gigabyte of information readily available, and thus, a content analysis of the internet could provide answers to numerous research questions.

**Data and Research Reports Need to Be Accessible**

Systematic sharing of data and dissemination of research reports needs to be institutionalized. There are several reasons for advancing this agenda item. As already noted, many published content analyses do not provide the detail needed to evaluate or replicate the work or to compare it with other coding systems. Making the CHAMP coding system public and readily available will provide the research community with a resource for replication and extension. To facilitate such activity, we also plan to make public the coding data upon which our analyses are drawn. This will permit others to
conduct novel analyses using the same database and to compare results with those already conducted. Finally, providing a website where comprehensive trends in risk and protective behaviors are cataloged will help to reduce redundant efforts by researchers who might otherwise be unaware of the findings and encourage those in the field to examine novel questions not answered by existing analyses. One example of a website that has similar objectives is directed by Michael Rich at his Center on Media and Child Health at Harvard. This site contains a database of existing literature on media and child health that is available free on the web: http://cmch.tv/research/searchCitations.asp.

The Annenberg Coding of Health and Media Project (ACHAMP) intends to provide a one-stop shop for locating content analyses of the identified behaviors. The ACHAMP will also distribute reports on health and well-being as seen in the media and will encourage others to conduct their own analyses using the coding data derived from the project.

**Conclusion**

At best, past work, including our own, has provided only a collection of snapshots, not a panoramic view of the portrayals of health behavior in the media. The problems with past studies that we outlined here limit our collective knowledge on this important subject. Perhaps unfairly, we have shone a light on limitations without providing comparable discussion of the many positive aspects of these studies. The advances achieved by past research in media content analysis deserve their own volume. We do not wish to minimize the importance of these advances. Instead, we wish to highlight opportunities for future progress. Taken together past literature does not paint a comprehensive picture of health behavior in media. Our intention is to call for a greater
integration and a comprehensive response to these limitations. The Annenberg Coding of Health and Media Project (ACHAMP) aims to do just that.
References


