Political Communication, 34:323–343, 2017

ISSN: 1058-4609 print / 1091-7675 online
DOI: 10.1080/10584609.2016.1235640

Routledge
Taylor & Francis Group

How Measurement Error in Content Analysis and Self-Reported Media Use Leads to Minimal Media Effect Findings in Linkage Analyses: A Simulation Study

MICHAEL SCHARKOW and MARKO BACHL

In the debate on minimal media effects and their causes, methodological concerns about measurement are rarely discussed. We argue that even in state-of-the-art media-effects studies that combine measures of media messages and media use (i.e., linkage analyses), measurement error in both the media content analysis and the media use self-reports will typically lead to severely downward-biased effect estimates. We demonstrate this phenomenon using a large Monte Carlo simulation with varying parameters of the content analysis and the survey study. Results show that measurement error in the content analysis and media use variables does indeed lead to smaller effect estimates, especially when the media messages of interest are relatively rare. We discuss these findings as well as possible remedies and implications for future research.

Keywords content analysis, linkage analysis, media effects, media use, Monte Carlo simulation, reliability

Media effects in political communication and particularly during election campaigns have often been found to be rather small. Consequently, eras of minimal effects of mass media have repeatedly been diagnosed, first following the seminal work of Lazarsfeld, Berelson, and Gaudet (1968), but also in the light of more recent developments of media fragmentation and polarization (Bennett & Iyengar, 2008). Theoretical and methodological developments have contributed to a more differentiated state of research beyond the minimal effects paradigm (Iyengar & Simon, 2000). Yet a general dissatisfaction remains: It is highly counterintuitive that empirically established media effects tend to be small even though almost all contacts with politics and politicians in modern societies happen via mass media. This obvious contradiction led some scholars to the assessment that the “state of research on media effects is one of the most notable embarrassments of modern social science” (Bartels, 1993, p. 247).

In addition to theoretical explanations for the minimal effects findings, researchers have rethought their methodological approaches in order to better detect media effects. The

Michael Scharkow and Marko Bachl are Postdoctoral Researchers, Department of Communication, University of Hohenheim

Address correspondence to Michael Scharkow, University of Hohenheim, Department of Communication, Fruwirthstraße 49, 70599 Stuttgart, Germany. E-mail: michael.scharkow@uni-hohenheim.de

reliable measurement of exposure to relevant messages was identified as a major concern in typical survey studies. Maybe most prominently, Price and Zaller (1990, 1993) argued that media use self-reports in surveys have only little value for predicting interpersonal differences in exposure to relevant political messages. This led, for one, to a surge of experimental studies in political communication research (Arceneaux, 2010; Iyengar, 2001, 2011). In controlled experimental settings, message exposure is measured without error and larger effects of political stimuli are demonstrated. However, the external validity of experimental results is still limited, all improvements of online methods notwithstanding. The criticism remains that the larger effects are not only caused by better measurement, but also by the artificial experimental context.

Another methodological innovation that addresses the validity concerns of earlier observational survey studies is to approximate each individual respondent's exposure to relevant media messages instead of using only media use measures. A message exposure score is constructed by linking media message measures from a content analysis to the self-reported media use of each individual respondent (hereafter: *linkage analysis*).[1] The approach has been recognized as the "state-of-the art analysis of the impact of specific news consumption on political behavior" (Fazekas & Larsen, 2016, p. 196) even in critical assessments.

In this article, we address the consequences of an important yet rarely discussed challenge for detecting media effects in linkage analyses: Measurement error that occurs both in the content analysis and in the survey parts of the studies. Based on a review of published linkage analyses and the literature on the quality of media use self-reports and content analytical measures, we argue that measurement error from both sources may lead to severely downward-biased effect estimates, erroneously suggesting minimal media effects. The consequences of measurement error are evaluated with a large-scale Monte Carlo simulation. We show that measurement error in both parts of a study indeed deflates the effect estimates, particularly under the condition of low prevalence of the relevant media messages. We conclude by discussing these findings as well as possible remedies and implications for future methodological and substantial research.

Introduction to Linkage Analysis

The idea of constructing individual measures of exposure to relevant media messages was first explicitly described and implemented by Miller, Goldenberg, and Erbring (1979) with data from the 1974 American National Election Study (see also Erbring, Goldenberg, & Miller, 1980). In their provocatively written section "Linkage Analysis (or, Not All Media Content Is Alike)", Miller and colleagues (1979, p. 68) laid out the main motivation for combining content analyses and surveys at the individual level:

> "Directional," evaluative media research has been confounded by the frequent use of, at best, indirect measures of media impact which require large inferential leaps. It has suffered from the methodological problems inherent in a failure to distinguish between reliance on a medium and exposure to a message, or between exposure in general and exposure to particular message content.[…] Only an interconnected data set of survey responses and media content allows one to move beyond analysis based on measures of media exposure or media message alone to consider the actual media content to which people have been exposed.

Similar arguments have been reiterated some decades later in discussion pieces on the state of media-effects research methods (e.g., Schuck, Vliegenthart, & de Vreese, 2016a; Shoemaker & Reese, 1990; Slater, 2004, 2016; Valkenburg & Peter, 2013). Instead of predicting some outcome of interest by media *use* and implicitly assuming what kind of media *messages* the respondents were exposed to, researchers should explicitly take the media messages into account.

Linkage analysis proved to be a useful research design across a range of research topics in political communication. Early on, it was implemented in several agenda-setting studies (Erbring et al., 1980; Hügel, Degenhardt, & Weiss, 1989; Rössler, 1999). Other pioneering studies include Dalton, Beck, and Huckfeldt's (1998) analysis of media effects in the 1992 U.S. presidential elections, research on mediatized conflicts (Kepplinger, Brosius, & Staab, 1991), research on the effects of news about unemployment (Mutz, 1992), and research on the role of news complexity in knowledge gap research (Kleinnijenhuis, 1991).

Today, linkage analyses are common in political media-effects research, many of them conducted in the context of European election campaigns. The combination of content analyses and surveys was used to explain the perception of the European Union (Peter, 2003), the vote in a Danish referendum campaign (de Vreese & Semetko, 2004), the vote for Euroskeptical parties (van Spanje & de Vreese, 2014), and the importance of party leader evaluations for vote choice (Takens, Kleinnijenhuis, Hoof, & Atteveldt, 2015). Others studied the effect of strategy news on political cynicism (Schuck, Boomgaarden, & de Vreese, 2013) and the effect of the availability of relevant information in the media (Elenbaas, Boomgaarden, Schuck, & de Vreese, 2013) or infotainment (Jebril, de Vreese, Dalen, & Albæk, 2013) on political knowledge. Swiss referendum campaigns were the contexts for studies of media effects on stereotypic attitudes toward immigrants (Schemer, 2012) and on the timing of voting decisions (Matthes, 2012). Kepplinger, Geiss, and Siebert (2012) used a linkage analysis for the detection of framing effects during scandals.[2]

All linkage analyses follow the same basic logic (a worked example of the technical procedures is given in the Technical Appendix provided in the online supplemental material):

1. Media messages that are theoretically assumed to affect the outcome of interest are measured in a content analysis (Schuck et al., 2016a). The measurement of theoretically relevant media messages might be as simple as coding whether a topic is present in a news story (e.g., in agenda-setting studies; Erbring et al., 1980; Rössler, 1999). Other one-item measures include the valence of portrayals of politics and politicians (Desmet, van Spanje, & de Vreese, 2015; Miller et al., 1979) or the position of arguments (pro or contra) about an issue that is subject to a referendum (Matthes, 2012; Schemer, 2012). More complex media messages such as strategic frames (Elenbaas & de Vreese, 2008; Schuck et al., 2013) or scandal frames (Kepplinger et al., 2012) are often measured by coding multiple indicators. The result of Step 1 is a content analysis data set that contains at least the message variables of theoretical interest and the media outlet for each coding unit (i.e., news item, sentence, or any other unit for which the content characteristics are recorded). Further characteristics, such as publication date or prominence, may also be recorded if the panel survey (Step 3) contains more than two waves or more fine-grained weighting procedures are applied in Steps 2 or 4.
2. The media message variables are aggregated for each media outlet and, possibly, time period.[3] Simple summaries like the (relative) frequency of news items with the relevant message or the difference of the number of news items with positive and negative

evaluations are used in most studies. Some studies implemented more complicated measures, for example by accounting for the prominence of the news items (Kepplinger et al., 2012; Peter, 2003) or by applying a higher weight to negative news (Boomgaarden, van Spanje, Vliegenthart, & de Vreese, 2011). The result of Step 2 is a data set of the aggregated media message variables, often as (weighted) counts or proportions (of items with the relevant message).

3. The survey part of a linkage analysis contains measures of the outcomes of interest, measures of self-reported media use, and other theoretically important predictors of the outcome for each individual respondent. Most earlier studies relied on cross-sectional surveys. Modern linkage analyses routinely use panel designs to strengthen causal inference. Most studies build on two-wave panels, but designs with up to 11 waves can be found in the literature (Takens et al., 2015). Self-reported media use is most often measured by asking the respondents about their usage frequency (days in the [past] week) of a set of media outlets. Other versions are open-ended or multiple-choice questions about the most important media source(s), the media outlet(s) used regularly, or the outlet(s) used yesterday. A measure of attention to the news is sometimes additionally included.
4. In the actual *linkage* step, each individual respondent is matched with the media message variables from Step 2 based on the media use that he or she reported in Step 3. The media message variables of all media outlets that were used by a respondent are summarized in a message exposure variable. The new variable approximates each individual's exposure to the media messages of theoretical interest (hereafter: *message exposure measure*). It is usually weighted by the individual usage frequency of each outlet, provided this information was recorded in Step 3. Some studies applied more complicated weights (for example, based on attention to the news [Elenbaas & de Vreese, 2008] or based on the temporal differences between the publishing dates of the news items and the dates of the survey interviews [Takens et al., 2015]).
5. Finally, the outcome of interest is regressed on the message exposure measure and other predictors. The regression weight of the message exposure measure is the quantity of interest for the effect of being exposed to the media messages that were measured in Steps 1 and 2.

Overall, linkage analyses contributed remarkably to the state of research on media effects of all kinds and particularly to our knowledge about media effects during political campaigns. Yet the effects revealed in these studies were comparatively small. One reason for this is discussed in the next section: measurement error that occurs in both parts of a linkage analysis.

Measurement Error in Media-Effects Research

Reliability and Validity of Media Use Measures

For at least the past four decades, communication scholars have investigated the reliability and validity of self-reported media use measures (Allen, 1981; Chaffee & Schleuder, 1986; Lee, Hornik, & Hennessy, 2008; Prior, 2009) and their consequences for media-effects research (Ansolabehere & Iyengar, 1995; Bartels, 1993; Romantan, Hornik, Price, Cappella, & Viswanath, 2008; Zaller, 2002). In political communication research, the update of the media use measures in the American National Election Studies (ANES) panel has recently revived the debate about the validity and reliability of such self-reports and alternative approaches to measuring political news use (Dilliplane, Goldman, & Mutz, 2013; Goldman, Mutz, & Dilliplane, 2013; LaCour & Vavreck, 2014; Prior, 2013).

Following the literature on measurement theory, one can distinguish between random and systematic error. Both can be found in media use measures, and both have consequences for studying media effects. Random errors in self-reported behavior are just as common in communication research as in other areas. They can be explained by many different processes, including respondents' satisficing behavior or limited recall capacity (Schwarz & Oyserman, 2001). Since media use is often only measured using single items, most studies have relied on panel data to test the reliability and stability of the measures. Results by Allen (1981), Bartels (1993) or, more recently, Lee and colleagues (2008) have shown that most media use measures, both of general media use and specific media outlets, have Wiley and Wiley (1970) reliabilities in the range of .6 to .8. Similarly, Dilliplane and colleagues (2013) reported Heise (1969) reliabilities between .52 and .88 for different measures of (political) TV use.

The second impairment of media use self-reports is systematic misreporting. Systematic misreporting, or more specifically, exaggeration of (news) media use, has been a known issue at least since the early 1970s (Allen, 1981, p. 235), and seems just as relevant today. Comparing self-reports with passive measures, frequent overreporting was recently found by Wonneberger and Irazoqui (2016) for general television use and by Scharkow (2016) for internet use. In the field of political communication, the problem of systematic misreporting was discussed in detail in the influential reports of Price and Zaller (1990, 1993). Prior (2009) used aggregates of passively tracked Nielsen data to estimate that self-reports exaggerate news use by a factor of 3 on average. LaCour and Vavreck's (2014) individual-level comparison of self-reports with a passively tracked measure showed that almost half of the respondents overreported their use of television news. In light of the findings on news media use, our simulation will focus on the consequences of overreporting, because it is considered as the predominant concern for measuring political media use.[4]

What are the expected consequences of random and systematic errors in media use measures? For purely random measurement error, the axioms of classical test theory hold: All bivariate correlations between media use and other variables will be attenuated, and media effects will be underestimated. The relationships in multivariate models can be downward- or upward-biased, but underestimation of the relationships is more common in most circumstances (Bartels, 1993). This consequence of imperfect measurement of media use has been known as an obstacle to detecting media effects for several decades (McGuire, 1986), yet little has been undertaken to study the phenomenon more deeply in the context of political communication (but see Bartels, 1993, for an exception).

Overreporting of media use has similar consequences as long as the measurement error is not related to the true media use score, with three additional caveats: First, overreporting will deflate the absolute estimate of the media use effect as quantified, for example, by the raw coefficient in a linear regression. If, as a simple example, all respondents reported twice the amount of news hours that they actually viewed, the estimated effect of an hour of news viewing would only be half as large as the true effect. However, a bias of the absolute estimate is not much of a concern in most studies, because we rarely have a precise assumption about its size in the first place. Second, as Zaller (2002) noted, strong overreporting not only affects the validity, but also the reliability of the media use measure because overreporting for a bounded response variable will decrease its variance compared to the true media use. Third, if respondents consistently exaggerate their responses, overreporting might increase the *estimated* reliability of a measure (Prior, 2013; Zaller, 2002). This can cause a false sense of trust in a measure that is in fact not reliable.

Reliability of Media Message Measures

Issues of reliability have been essential concerns in content analysis since its inception in the 1940s and 1950s. Almost every textbook contains at least one chapter about this topic, in which strategies for ensuring reliable coding and conducting reliability tests are treated in-depth (Krippendorff, 2013). But even with clear codebooks and extensive coder training, some measurement error will inevitably be present in the collected data due to textual or visual ambiguity of the coding units. The consequences of a misclassified media message variable for a linkage analysis are far from obvious. Increased problem awareness and revised publication standards have ensured that almost every recent linkage analysis reported results of a reliability test. Little attention has been devoted, however, to the question of how the imperfect coding affects the media effect estimate from these studies. We know only of a single explicit statement on the issue: "The imperfect reliability of the media content measure should attenuate media effects.[…] Because data are unavailable to estimate these measurement effects empirically, their precise influence can only be surmised" (Dalton et al., 1998, p. 122). In order to better understand the implications of unreliable media message variables for linkage analysis, it is again useful to distinguish between random and systematic error. In the context of content analysis, systematic error or misclassification means that some wrong categories are more likely to be coded than others given the true value of the coding unit. It is easy to see how systematic misclassification would affect the results of a linkage analysis: If news stories are systematically misclassified as containing a certain message when they actually do not, the prevalence of this message will be overestimated. The individual message exposure measure as computed in Step 4 of the linkage analysis (provided earlier) will in turn be inflated—that is, the exposure to the media message of interest will be exaggerated. Similar to the consequences of consistent overreporting of media use, overestimation (or underestimation) of the exposure to a message will bias only the raw effect estimate of message exposure, at least if the systematic misclassification is consistent across all media outlets. As noted earlier, we usually do not worry so much about the size of the raw effect estimate. Our simulation study will not take systematic coding error into account, because we have no reason to believe that systematic misclassification conditional on the media outlet is a widespread concern.

Random misclassification happens when coders mistakenly enter a wrong code regardless of the true value—that is, all wrong categories have the same probability of being coded. Random measurement error is often considered harmless for univariate analysis, because it is well-known that the point estimates of univariate summaries are unbiased by random normal error. This does not, however, apply to most content analyses, since for categorical variables, random measurement error can lead not only to attenuated or inflated correlations (Dosemeci, Wacholder, & Lubin, 1990), but also to severely biased point estimates of the category proportions (Schwartz, 1985). In short, in the presence of random misclassification, all proportion estimates will be biased towards $1/k$, where k is the number of categories. The estimated proportion of a variable that measures whether a message is present in a news story will be biased toward a prevalence of 50%. This problem becomes more severe with lower reliability and more extreme proportions. If, for example, the true proportion of stories with a message in the news is 5%, even a high-quality content analysis with a true-score reliability of .9 would yield an estimate of 14%. A still-acceptable coding with a reliability of .7 would result in a proportion estimate of 32%, more than 6 times the true message frequency.[5] The *differential* bias toward a uniform distribution as a function of random misclassification and true proportions can have severe implications for linkage analyses: If media outlets vary in the frequency of covering a message, random misclassification decreases the variance between media outlets by shifting outliers to the average, which in turn weakens the linkage analysis.The coding reliability and the prevalence of the messages of interest are important influences on the ability of a linkage analysis to detect the true media effect. Most modern linkage analyses reported reliability statistics. They ranged from values as low as Krippendorff's $\alpha \approx .6$ (e.g., Desmet et al., 2015; Schuck et al., 2013; Schuck, Vliegenthart, & de Vreese, 2016b) to satisfying values above .9. Our simulation thus considers a broad range of values from .5 to 1.[6] With regard to typical distributions of the content analysis data, we have less information, because descriptive statistics of the media message variables are only infrequently reported in published linkage analyses. There is, however, evidence that researchers are sometimes interested in the effects of the exposure to media messages with low prevalence. In a study of the effects of the news coverage about immigration and crime, the shares of news items with relevant topics ranged from 0.6% to 10.1% (Burscher, van Spanje, & de Vreese, 2015, Appendix A). Another study reported shares between 0% and approximately 16% for news items with the messages of interest (Jebril et al., 2013, Figure 1). More examples of studies with low prevalence of at least some of the messages of interest can be found (e.g., Kepplinger et al., 2012; Schuck et al., 2016b; van Spanje & de Vreese, 2014). We will therefore include proportions from .05 to .5 in our simulation study.

Conclusion of the Literature Review

Linkage analyses are well-suited to detect media effects in observational studies, because they promise to incorporate inter-individual differences in exposure to relevant media messages. But the combination of two data sources comes at a cost: The message exposure variable is also subject to two sources of measurement error. Previous work has demonstrated that a lack of reliable media use self-reports can and does lead to attenuated media effect estimates in political communication research (Ansolabehere & Iyengar, 1995; Bartels, 1993; Zaller, 2002). The implications of misclassification in the content analysis part have not yet been studied, but convincing reasons exist to assume that it also leads to underestimation of the true effects. The consequences should be more severe if the prevalence of the messages of interest is low.Since linkage analyses themselves are quite complex already, it is difficult to ascertain analytically the impact of measurement error from two sources in the multiplicative message exposure variable. Intuitively, the combination of two imperfectly measured variables should lead to an exacerbation of the overall measurement error, and, in turn, to downwards-biased effect estimates. In the remainder of this article, we investigate the effectiveness of a linkage analysis in the presence of measurement error by means of a large-scale Monte Carlo simulation.

Reliability of Media Use Self-Report (rows) × Reliability of Content Analysis (columns)

Overreporting: none

	Message Frequency: .05						Message Frequency: .10						Message Frequency: .25						Message Frequency: .50					
	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1
.5	.33	.34	.35	.37	.42	.51	.33	.35	.37	.41	.46	.51	.34	.38	.42	.46	.49	.51	.37	.42	.46	.49	.50	.50
.6	.40	.41	.42	.45	.50	.61	.40	.42	.45	.49	.55	.61	.41	.45	.50	.55	.59	.61	.45	.51	.55	.59	.61	.61
.7	.46	.48	.49	.52	.58	.71	.47	.50	.52	.57	.64	.71	.48	.53	.58	.64	.69	.71	.53	.59	.65	.69	.71	.71
.8	.53	.54	.56	.59	.66	.81	.53	.56	.60	.65	.73	.81	.54	.60	.66	.73	.79	.81	.60	.67	.74	.78	.80	.81
.9	.59	.60	.62	.66	.74	.90	.59	.63	.67	.73	.81	.90	.60	.67	.74	.81	.88	.90	.67	.75	.82	.87	.90	.90
1	.65	.67	.69	.73	.82	1	.66	.69	.73	.80	.90	1	.67	.74	.82	.90	.97	1	.74	.83	.91	.96	.99	1

Overreporting: 1 day

	Message Frequency: .05						Message Frequency: .10						Message Frequency: .25						Message Frequency: .50					
	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1
.5	.32	.32	.33	.36	.40	.49	.32	.34	.36	.39	.44	.49	.33	.36	.40	.44	.47	.49	.36	.41	.44	.47	.48	.49
.6	.39	.39	.41	.43	.49	.59	.39	.41	.43	.47	.53	.59	.40	.44	.48	.53	.57	.59	.44	.49	.54	.57	.58	.59
.7	.45	.46	.48	.51	.57	.69	.46	.48	.51	.55	.62	.69	.46	.50	.57	.62	.67	.69	.51	.58	.63	.66	.68	.69
.8	.51	.53	.54	.58	.65	.79	.52	.55	.58	.63	.71	.79	.53	.58	.64	.71	.76	.79	.58	.66	.72	.76	.78	.79
.9	.57	.59	.61	.64	.72	.88	.59	.61	.65	.71	.79	.88	.59	.65	.73	.79	.85	.88	.65	.74	.80	.85	.87	.88
1	.65	.66	.68	.73	.81	.99	.65	.69	.73	.79	.89	.99	.65	.73	.81	.89	.96	.99	.73	.82	.90	.95	.98	.99

Overreporting: 2 days

	Message Frequency: .05						Message Frequency: .10						Message Frequency: .25						Message Frequency: .50					
	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1
.5	.29	.30	.31	.33	.37	.45	.30	.31	.33	.36	.41	.45	.30	.33	.37	.41	.44	.45	.33	.38	.41	.43	.45	.45
.6	.36	.36	.38	.40	.45	.55	.37	.38	.40	.44	.49	.55	.37	.41	.45	.49	.53	.55	.40	.46	.50	.53	.54	.55
.7	.42	.43	.44	.47	.53	.64	.43	.44	.47	.51	.58	.64	.43	.47	.53	.58	.62	.64	.47	.54	.59	.62	.64	.64
.8	.49	.50	.51	.55	.61	.74	.49	.52	.54	.59	.67	.74	.49	.55	.61	.66	.72	.74	.55	.62	.67	.71	.73	.74
.9	.54	.56	.58	.61	.69	.84	.55	.58	.61	.67	.76	.84	.56	.62	.69	.75	.81	.84	.62	.70	.76	.81	.83	.84
1	.62	.63	.67	.70	.78	.95	.64	.66	.70	.76	.86	.95	.64	.70	.78	.86	.92	.95	.70	.79	.87	.92	.94	.95

Overreporting: 3 days

	Message Frequency: .05						Message Frequency: .10						Message Frequency: .25						Message Frequency: .50					
	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1
.5	.25	.26	.27	.29	.32	.39	.26	.27	.29	.32	.35	.39	.26	.29	.32	.35	.38	.39	.29	.33	.36	.38	.39	.39
.6	.31	.32	.33	.35	.39	.48	.31	.33	.35	.38	.43	.48	.32	.35	.39	.43	.46	.48	.35	.40	.44	.46	.47	.48
.7	.37	.38	.39	.42	.47	.57	.37	.39	.42	.45	.51	.57	.38	.42	.46	.51	.55	.57	.42	.47	.52	.55	.56	.57
.8	.44	.44	.46	.49	.55	.66	.43	.46	.49	.53	.60	.66	.44	.49	.54	.60	.64	.66	.49	.55	.60	.64	.66	.66
.9	.49	.51	.53	.56	.63	.76	.50	.53	.56	.62	.69	.77	.51	.57	.63	.69	.74	.76	.56	.64	.70	.74	.76	.77
1	.58	.59	.62	.66	.73	.89	.59	.62	.66	.71	.80	.89	.60	.66	.73	.80	.86	.89	.66	.74	.81	.86	.88	.89

Figure 1. Simulation results: True-score reliability of the observed message exposure. *Notes*. The figure presents the true-score reliabilities $Rel_{Message\ Exposure}$ as a function of the reliability of the content analysis Rel_{Coding} on the x-axis, the reliability of the media use self-reports $Rel_{Media\ Use}$ on the y-axis, the proportion of the news items with the relevant message $p^{true}_{Message}$ in the horizontal facets, and the amount of overreporting $OR_{Media\ Use}$ in the vertical facets. Cell entries are the means of 1,000 simulation runs.

Methods

Monte Carlo simulation is a useful tool for evaluating the implications of changes of single components in complex models. As Zaller (2002) noted in his simulation study on the power of election studies to uncover media effects, the specification of a simulation "is partly guesswork, [but] it is not unconstrained guesswork" (p. 313). Choices have to be made with regard to the data-generating processes, the distributions of the involved variables, and the input parameters. The challenge is to specify a simulation that is simple enough to be comprehensible, yet realistically models the underlying process of interest. The specification of the present simulation is informed by a thorough review of published linkage analyses and the literature on measurement error in media use self-reports and content analyses. Most linkage analyses followed the general steps outlined earlier, but no two studies were alike in every detail. We thus implemented a design that might be best described as a simplified version of the most common modern linkage analysis. The focus of the simulation study is to explore and quantify the implications of measurement error in media use self-reports and media message variables for detecting the effect of being exposed to certain media messages. Earlier we gave our reasons for which types of measurement error will be considered and which will not. A linkage analysis involves many more decisions by the researchers that cannot be part of the present simulation.

Design

A linkage analysis in the Monte Carlo simulation followed the five steps that were described earlier for their empirical counterparts. The notable difference is that both the true data and the observed data (with error) were generated or used, respectively, in each step. One run of the simulation proceeded as follows:[7]

1. We simulated the *true* media coverage of 10 outlets that published 30 news items each during the study period. The overall occurrence of the relevant media messages was set to the proportion $p^{true}_{Message}$. The proportions were set to vary across the news outlets, such that some outlets provided more relevant messages than others. The *observed* media coverage was then generated by simulating a coding process with random misclassification and a true-score reliability of Rel_{Coding}. The values of $p^{true}_{Message}$ and Rel_{Coding} were varied in each run.
2. The media message variable was calculated for each outlet as the sum of news items with the message of interest, both for the *true* and the *observed* media coverage.
3. We simulated the *true* data from a two-wave panel survey with 500 respondents using 10 media outlets. Each respondent was randomly assigned an initial value for the outcome variable from a standard normal distribution. For each media outlet and each respondent, the true media use was randomly sampled from discrete values in a range from 0 to 7, representing the typical self-report question about on how many days of a week an outlet has been used by a respondent. We then added measurement error to the true value to get the *observed* media use measure for each media outlet and each respondent. We considered two types of measurement error: Random measurement error was added in order to create an observed media use measure with a true-score reliability of $Rel_{Media\ Use}$. Systematic overreporting was simulated by shifting the mean of the true media use measure by a value of $OR_{Media\ Use}$. The values of $Rel_{Media\ Use}$ and $OR_{Media\ Use}$ were varied in each run.
4. The message exposure variable was computed by weighting the media message variables from Step 2 by the media use variables from Step 3. We also calculated and saved the true-score reliability of the message exposure measure, $Rel_{Message\ Exposure}$.
5. We assumed a simple linear media-effects model where the outcome at the time of the second panel survey y_t is a function of the prior outcome y_{t-1}, the true message exposure x^{true}, and some random error ϵ. The process is described by the panel regression

$$y_{\text{t}} = \beta_1 * y_{t-1} + \beta_2^{\text{true}} * x^{\text{true}} + \epsilon,$$

where β_1 quantifies the effect of the lagged outcome variable, and β_2^{true} quantifies the true media effect.

We specified the regression weights in units of standard deviations of the predictors and the outcome (standardized effects), because there is no information on what absolute effect to expect from being exposed to one additional message. We assumed rather high stability of the outcome with $\beta_1 = .8$, because it is well-known that political attitudes and other relevant outcomes in political communication research are hard to change even during campaigns. The *true* media effect was set to a substantial yet realistic quantity of $\beta_2^{\text{true}} = .2$. The combination of the regression weights yielded a residual standard error of $\varepsilon \approx 0.6$ and an adjusted $R^2 \approx .7$, which can be considered realistic for a panel model that included the lagged outcome variable as predictor. Finally, we regressed the outcome variable $\mathrm{y_t}$ on the prior outcome y_{t-1} and the observed message exposure score x^{obs}. The regression weight β_2^{obs}, which quantified the *observed* media effect, was saved for every run of the simulation.

Simulation Parameters and Outcome Measures

Table 1 summarizes the input parameters and the outcomes. The final Monte Carlo simulation used a 4 ($p_{Message}^{true}$) x 6 (Rel_{Coding}) x 6 ($Rel_{Media\ Use}$) x 4 ($OR_{Media\ Use}$) full factorial design with 1,000 replications and therefore yielded 576,000 results, which are described and analyzed next. The outcome of interest is the ratio $\beta_2^{\text{obs}}/\beta_2^{\text{true}}$ (hereafter: *relative bias*). This coefficient can easily be interpreted as the proportion of the true effect that is uncovered in the linkage analysis. Values larger than 1 indicate inflation and values smaller than 1 indicate attenuation. As a by-product of the simulation study, we also present the true-score reliability of the combined message exposure variable, $Rel_{Message\ Exposure}$ from Step 4. This measure might prove useful for researchers interested in using errors-in-variables modeling that often requires reliability estimates of single variables (Hardin & Carroll, 2003).

Results

The results for all conditions of the Monte Carlo simulation are summarized in Figures 1 and 2. The reliability of the observed message exposure measure and the relative bias of the observed media effect, respectively, are presented as a function of the reliability of the content analysis, the proportion of relevant news items, the reliability of the media use self-reports, and the overreporting of media use. Each facet shows a combination of a level

Table 1
Input parameters and outcomes of the Monte Carlo simulation

Symbol	Description	Simulated Values
Input parameters		
$p_{Message}^{true}$	Proportion of news items with a relevant message	.05, .10, .25, .50
Rel_{Coding}	Reliability of the content analysis	.5, .6, .7, .8, .9, 1.0
$Rel_{Media\ Use}$	Reliability of the media use self-reports	.5, .6, .7, .8, .9, 1.0
$OR_{Media\ Use}$	Overreporting of media use variable in days	0, 1, 2, 3
Outcomes		
$Rel_{Message\ Exposure}$	Reliability of the message exposure variable	
Relative Bias	Ratio of the observed to the true media effect	

		Message Frequency: .05						Message Frequency: .10						Message Frequency: .25						Message Frequency: .50					
Overreporting	Reliability of Media Use Self-Report \ Reliability of Content Analysis	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1	.5	.6	.7	.8	.9	1
Overreporting: none	.5	.58	.58	.59	.61	.64	.71	.57	.59	.60	.63	.68	.71	.58	.61	.64	.68	.71	.71	.61	.65	.67	.69	.71	.71
	.6	.63	.64	.65	.67	.70	.78	.63	.65	.67	.70	.74	.78	.63	.67	.71	.75	.77	.78	.67	.71	.74	.76	.78	.78
	.7	.68	.69	.70	.73	.77	.84	.69	.70	.72	.75	.80	.84	.69	.73	.76	.80	.83	.84	.73	.77	.80	.83	.84	.84
	.8	.72	.73	.75	.77	.82	.90	.73	.75	.77	.80	.86	.90	.73	.77	.81	.85	.88	.90	.77	.82	.86	.89	.89	.90
	.9	.77	.77	.79	.81	.86	.95	.77	.78	.82	.85	.90	.95	.77	.81	.86	.90	.94	.95	.82	.86	.91	.93	.95	.95
	1	.80	.81	.83	.85	.91	1	.81	.83	.86	.89	.95	1	.82	.86	.90	.95	.98	1	.86	.91	.96	.98	1	1
Overreporting: 1 day	.5	.56	.56	.58	.60	.63	.70	.57	.58	.60	.62	.66	.70	.57	.60	.63	.66	.69	.70	.60	.64	.67	.68	.69	.70
	.6	.63	.63	.63	.66	.70	.77	.63	.64	.65	.69	.73	.76	.63	.66	.69	.73	.76	.77	.66	.70	.73	.75	.77	.77
	.7	.67	.67	.69	.71	.75	.83	.68	.69	.71	.74	.79	.83	.68	.71	.75	.79	.82	.83	.71	.76	.79	.81	.82	.83
	.8	.72	.72	.73	.76	.80	.89	.72	.74	.76	.80	.84	.89	.72	.76	.80	.84	.87	.88	.76	.81	.84	.87	.89	.89
	.9	.75	.77	.78	.80	.85	.94	.76	.78	.81	.84	.89	.94	.76	.80	.85	.89	.93	.94	.81	.86	.90	.92	.93	.94
	1	.80	.81	.82	.85	.90	.99	.81	.83	.86	.89	.94	.99	.81	.85	.90	.94	.98	.99	.85	.91	.95	.97	.99	.99
Overreporting: 2 days	.5	.53	.55	.55	.58	.61	.68	.55	.56	.58	.60	.64	.67	.55	.57	.61	.64	.66	.66	.57	.62	.65	.65	.67	.67
	.6	.60	.60	.62	.64	.66	.74	.60	.61	.63	.66	.70	.74	.61	.64	.67	.70	.73	.74	.64	.67	.70	.72	.74	.75
	.7	.64	.65	.67	.69	.72	.80	.65	.67	.68	.72	.76	.80	.65	.69	.73	.76	.79	.80	.69	.73	.77	.79	.80	.80
	.8	.69	.71	.72	.74	.78	.86	.70	.71	.74	.77	.82	.86	.70	.74	.78	.82	.85	.86	.74	.79	.82	.84	.86	.86
	.9	.74	.75	.76	.78	.83	.92	.74	.76	.78	.82	.87	.92	.75	.78	.83	.87	.90	.92	.78	.84	.87	.90	.91	.91
	1	.78	.80	.82	.84	.89	.98	.80	.81	.84	.87	.93	.98	.80	.84	.88	.93	.96	.98	.84	.89	.93	.96	.97	.97
Overreporting: 3 days	.5	.50	.51	.51	.53	.57	.62	.50	.51	.53	.56	.59	.62	.51	.54	.56	.59	.62	.63	.53	.57	.60	.61	.63	.63
	.6	.55	.56	.57	.59	.63	.69	.56	.58	.58	.62	.66	.69	.56	.60	.62	.66	.68	.69	.59	.63	.66	.68	.69	.69
	.7	.60	.61	.62	.64	.68	.76	.61	.62	.64	.68	.72	.75	.61	.65	.68	.72	.74	.75	.64	.69	.72	.74	.75	.75
	.8	.66	.66	.68	.70	.74	.81	.65	.68	.70	.73	.77	.81	.66	.70	.74	.77	.81	.81	.70	.74	.78	.80	.81	.81
	.9	.70	.71	.73	.75	.79	.87	.71	.73	.75	.78	.83	.88	.71	.75	.79	.83	.86	.88	.75	.80	.83	.86	.87	.88
	1	.76	.77	.79	.80	.86	.94	.77	.78	.80	.84	.90	.94	.77	.81	.85	.89	.93	.94	.81	.86	.90	.92	.94	.94

Figure 2. Simulation results: Relative bias of the observed media effect. *Notes*. The figure presents the relative bias ($\beta_2^{obs}/\beta_2^{true}$) of the observed media effect as a function of the reliability of the content analysis Rel_{Coding} on the x-axis, the reliability of the media use self-reports $Rel_{Media\ Use}$ on the y-axis, the proportion of the news items with the relevant message $p_{Message}^{true}$ in the horizontal facets, and the amount of overreporting $OR_{Media\ Use}$ in the vertical facets. Cell entries are the means of 1,000 simulation runs.

of the message frequency and a level of the overreporting parameter. Within each facet, the levels of the reliability of the media use measure are in the rows and the levels of the reliability of the content analysis are in the columns. The input values of the reliability parameters increase from the left to the right and from the top to the bottom in each facet. Each cell entry is the mean of the respective outcome in 1,000 simulation runs.

As a model check, we first direct our attention to the bottom right cells in the first row of facets, where we find the results of the simulation under ideal conditions. As expected, the true-score reliabilities of the observed message exposure measure were perfect (Figure 1) and the observed media effect was equal to the true effect (Figure 2), when both the content analysis and the self-report measures were perfectly reliable and unbiased.

Moreover, Figure 2 reveals that linkage analyses are—under the conditions tested in this study—far more likely to underestimate media effects than to overestimate them: None of the conditions showed an average relative bias greater than 1. Overall, only 1.84% of all simulations produced a larger observed than true effect.

The well-known attenuating impact of unreliable media use self-reports is found by comparing the cell entries in the right-most columns in the first row of facets from top to bottom. Assuming a perfect content analysis and no systematic overreporting of media use, the reliability of the message exposure variable was approximately equal to the reliability of the media use self-reports (Figure 1, top panel), and the observed media effect was attenuated in the presence of measurement error in the self-reports. For example, assuming a high reliability of the self-report of $Rel_{Media\ Use}$ = .8 and otherwise perfect measurement, the linkage analysis uncovered 90% of the true media effect (Figure 2, top panel, right-most column).

The effect of the second survey-related parameter, overreporting of media use, can be seen by comparing the facets from top to bottom. Overreporting overall caused only little relative bias in the observed media effect, which is in part due to the fact that we look at the effects in units of standard deviations. Comparing the top and bottom panels of Figure 2, we can see that 94% of the true media effect was uncovered if respondents severely overreported their media use by three days per week on average and otherwise perfect measurement.

The pattern of results in Figure 2 indicates that the relative bias of the observed media effect also varied substantially as a function of the content analysis parameters of the simulation, which is observable by comparing the cells in each block from left to right. Since the compact presentation of Figure 2 does not lend itself well for an in-depth review, we will proceed with a more detailed illustration of the results for selected typical conditions. Figure 3 displays the relative bias of the observed effect as a function of the reliability of the content analysis (x-axis), the reliability of the media use self-reports (horizontal facets), the overreporting in media use (vertical facets), and the proportion of news items with relevant messages (line types).

All four panels show the same pattern regarding the effects of random measurement error: A decrease in coding as well as self-report reliability led to a substantial decrease in the estimated media effect. Figure 3 furthermore reveals the nonlinear interaction of content analysis reliability and message frequency: The negative consequences of unreliable content analyses became worse when the messages of interest were rare. Studies with similar reliability in the content analysis and the survey measures can differ dramatically in the attenuation of the observed effect, depending on the message frequency—scholars who are interested in the effects of media messages with lower prevalence will *ceteris paribus* find smaller effects than those who analyze the effects of more common media messages.

From a communication researcher's perspective, it is relevant to know the relative magnitude of the negative effects stemming from various sources of measurement error. Specifically, one may ask, "Which reliability issues matter most?" One simple way of assessing this problem is to look at the relative difference in the observed media effects when one measure becomes less reliable while all others remain constant. Table 2 summarizes the relative bias in the observed media effect for various typical levels of reliability. We can see that, for a message frequency of $p^{true}_{Message} = .25$, a shift from perfect reliability in the media use self-reports to a typical level of .7 reduced the observed media effect to 84% of the true effect. The same loss of reliability in the content analysis led to a reduction to 90%. Combined, the two types of random measurement error reduced the observed effect to 76% of its true value. For relatively frequent messages, the

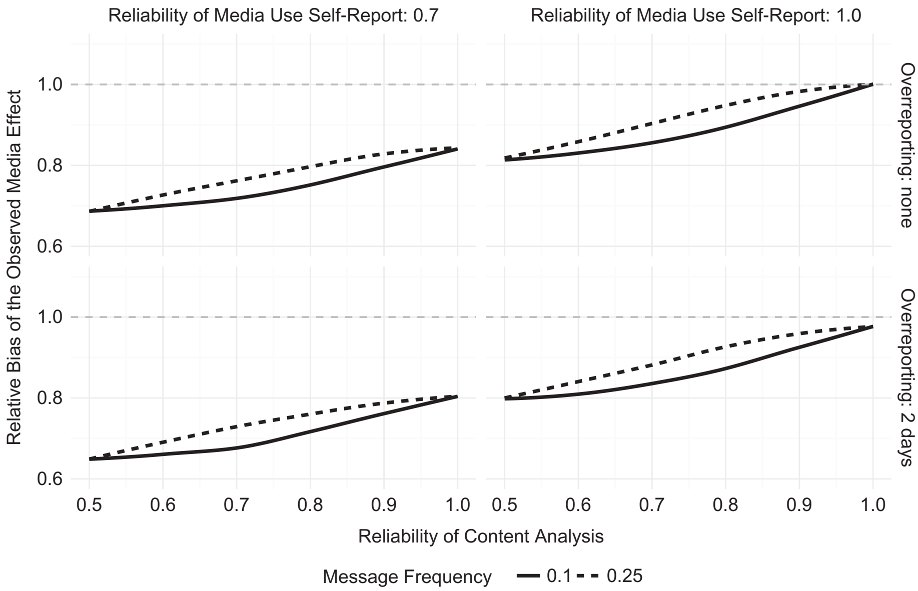

Figure 3. Relative bias of the observed media effect for selected conditions.

Table 2
Reliability of the observed message exposure measure and relative bias of the observed media effect at typical levels of measurement error and message frequency

Input Parameters			Outcomes	
$p^{true}_{Message}$	Rel_{Coding}	$Rel_{Media\ Use}$	$Rel_{Message\ Exposure}$	Relative Bias
0.05	0.7	0.7	0.49	0.70
0.05	0.7	1.0	0.69	0.83
0.05	1.0	0.7	0.71	0.84
0.05	1.0	1.0	1.00	1.00
0.10	0.7	0.7	0.52	0.72
0.10	0.7	1.0	0.73	0.86
0.10	1.0	0.7	0.71	0.84
0.10	1.0	1.0	1.00	1.00
0.25	0.7	0.7	0.58	0.76
0.25	0.7	1.0	0.82	0.90
0.25	1.0	0.7	0.71	0.84
0.25	1.0	1.0	1.00	1.00

Note. The table presents the results for selected combinations of message frequency and reliabilities of the content analysis and media use self-reports and no overreporting. The values for $Rel_{Message\ Exposure}$ and Relative Bias ($= \beta_2^{\text{obs}}/\beta_2^{\text{true}}$) are means from 1,000 simulation runs.

measurement error in the media use self-reports had more negative consequences than coding errors. Because of the interaction between message frequency and coding errors, things are different if we look at very rare messages ($p^{true}_{Message} = .05$). In this case, both content misclassification and errors in self-reported media use had similar consequences for the observed media effect.

Table 2 also displays the reliability of the combined message exposure variable for typical levels of measurement error in the content analysis and media use variables. Depending on the message frequency in the news, the reliability of the message exposure variable was as low as .49 for very infrequent media messages, assuming commonly reported reliabilities of .7 for the content and survey measures. Even when the message of interest was more frequent, the combined message exposure measure had a low reliability of .58, which in turn led to an underestimation of the true media effect by about 25%. There is, however, also a comforting result for studies with rather low reliability estimates for the content analysis: If the topic of interest is relatively common, the aggregation of the single codings into the media content measures at the level of media outlets in Step 2 of the linkage analysis will typically enhance the reliability. In the third row from the bottom of Table 2, it can be seen that if the survey measure was perfect and the message occurred in one-quarter of all stories, a coding reliability of .70 translated to a nominally better reliability of .82 for the message exposure measure.

Discussion

In this article, we contributed in two ways to the study of media effects in political communication. First, the summary of linkage analyses and the derivation of a formalized Monte Carlo simulation framework for their evaluation is in itself important. The systematization explicitly reveals the many major and minor decisions that are involved in designing a linkage analysis. Future simulation studies might build on and expand our model in order to better understand the complex relationships between characteristics of both the content analysis and the survey parts of the design, as well as the utility of statistical methods for their analysis. The Monte Carlo simulation framework furthermore provides researchers who plan or have already conducted a linkage analysis with a useful tool for a priori power analyses or post hoc sensitivity analyses. If the design of the linkage analysis is similar to the design described earlier, a convenient first step is to refer to the cells of Figures 1 and 2 that most closely fit the properties of the study. More detailed results can of course be derived by adapting the framework to the specifics of a study and carrying out simulations for a range of sensible values. For example, one could vary outlets systematically, creating a consonant or diverse media environment, or vary the simulated survey dates and time-specific media content, therefore creating temporal in addition to outlet-specific variation in the linkage analysis. Second, the results of the Monte Carlo simulation demonstrated that imperfect measurement in both the content analysis and the survey part of a linkage analysis led to a potentially severe underestimation of the media effect. It has been established before that unreliable media use self-reports are likely to attenuate media effect estimates in political communication research (Ansolabehere & Iyengar, 1995; Bartels, 1993; Zaller, 2002), and this implication is reflected in our results. Random measurement error in the self-reports alone deflated the observed effect to approximately 80% of the true effect for reliabilities at the lower bound of empirically tested measures. Systematic overreporting that is not correlated with actual media use turned out to be a relatively minor concern, at least as long as we are only interested in standardized effects. We note, however, that systematic misreporting will severely bias effect estimates that are quantified in the original units of the variables, such as the effect of being exposed to a certain amount of additional messages. We have not covered this issue in the present article, because the analytical reasoning is straightforward. Furthermore, we analyzed for the first time the consequences of measurement error in the content analysis for linkage analyses. The simulation revealed that misclassification during the coding process may also severely attenuate the observed media effect. Contrary to conventional wisdom, random, not systematic, misclassification of categories that are unevenly distributed is the culprit. The logic behind this finding is best explained by an illustrative example: Consider a very simple linkage analysis that includes only two media outlets, A and B. Five percent of the news stories in A and 30% in B contain the message of interest. A coding process with random misclassification and a true-score reliability of .9 produces observed proportions of 14% and 34% for A and B, respectively. Coding with worse but still-acceptable reliability of .7 yields observed proportions of 32% and 42%. Observe that not only the overall proportions increase, but also the difference between the outlets reduces significantly. In consequence, the individual message exposure variable that is constructed based on the misclassified media message measure will misrepresent the relative differences in exposure to the message between the respondents. Assuming the same number of news stories for A and B and the same usage frequency for all respondents, a user of A is in reality exposed to 6 times the amount of news items with the message compared to a user of B. The message exposure variable will instead estimate only a 2.4-fold or 1.3-fold dose of exposure. In short, random misclassification impairs the major advantage of a linkage analysis—that is, its ability to account for differences in message exposure—because media outlets are observed as being more similar in their relevant news coverage than they actually are. In sum, our Monte Carlo evidence suggests that measurement issues are at least in part responsible for minimal effect findings in political communication, even if researchers implement sophisticated and expensive linkage analyses. Even under the ideal conditions of a simulation study, measurement error of realistic amounts alone deflated the observed effects to sizes as low as 60% of the true effect. Empirical studies in realistic settings will quite likely recover even lower shares of the true effect. There is, however, also a positive note: The results suggest that media effects are, indeed, quite likely larger than previously reported. Given the assumptions of this simulation study, there is little evidence to suspect that media effects have been exaggerated in published linkage analyses. Most likely, the true effects are stronger than the observed effects, which is in line with the theoretical reasoning underlying most political communication research. Earlier research could only refer to the literature on the attenuating impact of unreliable survey measures and speculate on the consequences of coding errors in order to put the findings into perspective (see again the quote by Dalton et al., 1998, p. 122, provided earlier). We now add simulation evidence to support this reasoning. Finally, if we believe that minimal effect findings are at least in part caused by measurement error, we identify a problem that we as communication scholars can work on.

Recommendations

A recent review on the state of media-effects research concluded the following: “Overall, it seems safe to say that the progress of media-effects research will depend heavily on the improvement and further development of media exposure measures” (Valkenburg & Peter, 2013, p. 202). In light of our results, we are fully supportive of this conclusion. Linkage analysis already is a major improvement on survey-only designs, because the combination of survey data with content analysis data includes important additional information for detecting effects not only of media use, but also of message exposure (Miller et al., 1979; Schuck et al., 2016a). In order to uncover the true magnitude of media effects, more reliable measures and error-correction strategies need to be developed. In the context of linkage analyses, this concerns both survey and content analysis. While recent work primarily focuses on improving the measurement of media use through better self-reports in surveys, and alternatives such as media diaries or passive measures (see de Vreese & Neijens, 2016, for an overview), relatively little systematic research has focused on the content measures. As Shoemaker and Reese (1990) noted, "standardization of content analysis lags behind standardization in survey or experimental research" (p. 652). A first step to promote standardization would be the publication of codebooks and test material from reliability tests. Systematic methodological research on instruments for coding media messages comparable to scale development is desirable. Another promising approach to coding messages for linkage analyses is using computer algorithms in addition to or even instead of human coders. Automatic content analysis has several benefits for linkage analysis: First, the specific misclassification error probabilities of computer classification are often known or easy to estimate, and computer coding is highly deterministic. Second, one can easily employ an ensemble of multiple classifiers for the same message coding and increase the coding quality (Hillard, Purpura, & Wilkerson, 2008). Third, automatic coding can be used to classify huge amounts of media messages at little to no cost, which makes it attractive for linkage analyses that incorporate more than just a handful of frequently used outlets. We do not argue that automatic content analysis is inherently more reliable or valid than manual coding, but it offers benefits for linkage analyses that could even justify a moderate loss of reliability, especially when errors need to be accounted for anyway. Even with improved media use and content analysis measurement, some error will inevitably be present in the collected data, and this will bias the estimated media effects. Based on our simulation, most studies will substantially underestimate them. Put another way, scholars who detected media effects can be reasonably confident that a true effect existed and was larger rather than smaller. However, as communication scholars we should try to establish accurate estimates of media effects, and use the available data and tools for this purpose. Luckily, in most published linkage analyses, researchers already have a lot of information about measurement quality at their disposal: Reliability tests for the content analyses provide estimates about misclassification probabilities, and the reliability of the self-reports can either be estimated from the data (e.g., using the established approaches by Heise [1969] or Wiley and Wiley [1970]), or they can be reasonably derived from the recent literature on the reliability of media use. We therefore recommend to not only report such statistics, but use the information during data analysis. Most statistical software packages provide the means to incorporate measurement error into a regression model—for example, via latent variable models (Bollen, 1989; Fuller, 1987). Even with single-exposure scores, it is possible to fix the error variance of the score and obtain a corrected regression coefficient (for an example, see Hardin & Carroll, 2003), which should in most cases be larger than one obtained in a simple regression model and therefore worth the additional effort. Ultimately, we should investigate different methods that can deal with the different sources and types of error that can occur in a complex linkage analysis. For example, one could try to correct the content-analytic measures and self-reports separately before combining them into an exposure score. One flexible and powerful approach for error correction is simulation-extrapolation (Cook & Stefanski, 1994), which has been successfully applied to both survey and content analysis data (Benoit, Laver, & Mikhaylov, 2009), but not systematically evaluated for linkage analyses.

In sum, there are some effective and efficient ways to increase the reliability and validity of any linkage analysis. Therefore, we encourage scholars who plan a linkage analysis to (a) conduct meaningful reliability tests not only before, but also during, the actual coding phase with adequate sample sizes and state-of-the-art reliability measures that best approximate true-score reliability, (b) measure media use in every panel wave (ideally in at least three waves), so that the reliability of the self-reports can be empirically estimated using test-retest models, (c) estimate the reliability of the combined message exposure variable—for example, using our simulation, and (d) estimate a regression or path model in which the error variance of the message exposure variable (and all other variables, for that matter) is accounted for. Media effects in political communication might not be as minimal as they may appear in the light of measurement error in media message exposure. While linkage analysis already provides a valuable tool for measuring media effects, its usefulness depends on the reliability of the survey and media data that are linked. Therefore, we must continue to refine our measures and analytical approaches to further improve their quality and consequently their relevance for political communication research.

Notes

1. We chose the term *linkage analysis* to honor the contribution of Miller and colleagues (1979), who were—to our knowledge—the first to describe and implement this research design most similarly to the design of modern linkage analyses in the field of political communication. We acknowledge, however, that the idea to combine content analytical measures of media messages with media use reports from surveys has been implemented before. Notably, the seminal "People's Choice" study included "the index of political exposure bias," which weighted self-reports of exposure to several media items and campaign events with a measure of their leaning toward Republicans or Democrats (Lazarsfeld et al., 1968, pp. 177–178).
2. The examples intentionally exclude the line of research on the effects of political advertisements (mostly in the context of U.S. elections), which has much in common with media content linkage analysis (e.g., Franz & Ridout, 2007; Freedman & Goldstein, 1999) and shares some of its problems with regard to measuring exposure using self-reports (Vavreck, 2007). However, the measurement of the relevant media messages (i.e., the content of the ads) is comparatively straightforward and rich data sources exist to assist message measurement and geographic allocation of the ads (Goldstein & Freedman, 2002). In consequence, our main argument with regard to the combination of measurement error from *both* media content analysis and media use self-reports is less relevant. The results of the Monte Carlo simulation under the condition of (almost) perfect coding (i.e., $Rel_{Coding} \geq .9$) might give some indication for the ability of these studies to uncover true advertising effects. It should be noted, however, that the construction of individual measures of ad exposure also often differs from the procedures applied in mass media linkage analysis. Most importantly, information on ad buy has to be incorporated. Thus, the reliability of the ad exposure measure might differ from simple media use self-reports.
3. For the sake of simplicity, we assume only one media message variable and no temporal weighting for the following steps. Consequently, the subsequent description is limited to only one message exposure measure. If multiple relevant media messages are measured in Step 1, several message exposure measures can by constructed in Step 4 and used as predictors in Step 5. We also neglect the possibility of multistep aggregation, where, for example, media message variables are first aggregated from the coding units of single statement to the unit of news stories.
4. It should be noted that underreporting is also present in LaCour and Vavreck's (2014) data, mainly among those who reported never to watch TV news. Similarly, the amount of time spent watching TV is often underreported (Wonneberger & Irazoqui, 2016). However, most published

linkage analyses use frequency rather than duration measures, and are therefore more likely to encounter overreporting.

5. See the Technical Appendix in the online supplemental material for the formula to obtain the expected proportions under these assumptions.
6. The reported reliability of the media message variables depends on the choice of the reliability coefficient. Chance-corrected coefficients such as Krippendorff's α and Scott's π are currently the consensus measures for reliability in content analysis. Estimates based on these measures are (sometimes considerably) lower compared to simple measures of percent agreement and they depend on the marginal distributions in the test material. The message variable of a study that reports a percent agreement of .9 may well have a worse true-score reliability. In contrast, a message variable with an estimated reliability of Krippendorff's $\alpha = .6$ from a test with a highly skewed distribution of the message variable and a small sample may actually underestimate the true-score reliability. It is beyond the scope of the present article to join the debate about which coefficient is best under which circumstances. We merely note that media-effects researchers who want to evaluate the quality of their linkage analysis should follow the ongoing discussions about reliability measurement in content analysis and apply state-of-the-art measures that most closely approximate true-score reliability.
7. Details on the computation and estimation can be found in the Technical Appendix in the online supplemental material.

Supplemental Material

Supplemental data for this article can be accessed on the publisher's website at http://dx.doi.org/10.1080/10584609.2016.1235640.

References

Allen, R. L. (1981). The reliability and stability of television exposure. *Communication Research*, *8*(2), 233–256. doi:10.1177/009365028100800205

Ansolabehere, S., & Iyengar, S. (1995). *Messages forgotten: Misreporting in surveys and the bias toward minimal effects*. Retrieved from http://pcl.stanford.edu/common/docs/research/ansolabehere/1995/messages.pdf

Arceneaux, K. (2010). The benefits of experimental methods for the study of campaign effects. *Political Communication*, *27*(2), 199–215. doi:10.1080/10584601003709407

Bartels, L. M. (1993). Messages received: The political impact of media exposure. *American Political Science Review*, *87*(2), 267–285. doi:10.2307/2939040

Bennett, W. L., & Iyengar, S. (2008). A new era of minimal effects? The changing foundations of political communication. *Journal of Communication*, *58*(4), 707–731. doi:10.1111/j.1460-2466.2008.00410.x

Benoit, K., Laver, M., & Mikhaylov, S. (2009). Treating words as data with error: Uncertainty in text statements of policy positions. *American Journal of Political Science*, *53*(2), 495–513. doi:10.1111/j.1540-5907.2009.00383.x

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Boomgaarden, H. G., van Spanje, J., Vliegenthart, R., & de Vreese, C. H. (2011). Covering the crisis: Media coverage of the economic crisis and citizens' economic expectations. *Acta Politica*, *46*(4), 353–379. doi:10.1057/ap.2011.18

Burscher, B., van Spanje, J., & de Vreese, C. H. (2015). Owning the issues of crime and immigration: The relation between immigration and crime news and anti-immigrant voting in 11 countries. *Electoral Studies*, *38*, 59–69. doi:10.1016/j.electstud.2015.03.001

Chaffee, S. H., & Schleuder, J. (1986). Measurement and effects of attention to media news. *Human Communication Research*, *13*(1), 76–107. doi:10.1111/j.1468-2958.1986.tb00096.x

Cook, J. R., & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*(428), 1314–1328. doi:10.1080/01621459.1994.10476871

Dalton, R. J., Beck, P. A., & Huckfeldt, R. (1998). Partisan cues and the media: Information flows in the 1992 presidential election. *American Political Science Review*, *91*(1), 111–126. doi:10.2307/2585932

Desmet, P., van Spanje, J., & de Vreese, C. H. (2015). Discussing the democratic deficit: Effects of media and interpersonal communication on satisfaction with democracy in the European Union. *International Journal of Communication*, *9*, 3177–3198. doi:1932–8036/20150005

de Vreese, C. H., & Neijens, P. (2016). Measuring media exposure in a changing communications environment. *Communication Methods and Measures*, *10*(2–3), 69–80. doi:10.1080/19312458.2016.1150441

de Vreese, C. H., & Semetko, H. A. (2004). News matters: Influences on the vote in the Danish 2000 Euro referendum campaign. *European Journal of Political Research*, *43*(5), 699–722. doi:10.1111/j.0304-4130.2004.00171.x

Dilliplane, S., Goldman, S. K., & Mutz, D. C. (2013). Televised exposure to politics: New measures for a fragmented media environment. *American Journal of Political Science*, *57*(1), 236–248. doi:10.1111/j.1540-5907.2012.00600.x

Dosemeci, M., Wacholder, S., & Lubin, J. (1990). Does nondifferential misclassification of exposure always bias a true effect toward the null value? *American Journal of Epidemiology*, *132*(4), 746–748.

Elenbaas, M., Boomgaarden, H. G., Schuck, A. R. T., & de Vreese, C. H. (2013). The impact of media coverage and motivation on performance-relevant information. *Political Communication*, *30*(1), 1–16. doi:10.1080/10584609.2012.737411

Elenbaas, M., & de Vreese, C. H. (2008). The effects of strategic news on political cynicism and vote choice among young voters. *Journal of Communication*, *58*(3), 550–567. doi:10.1111/j.1460-2466.2008.00399.x

Erbring, L., Goldenberg, E. N., & Miller, A. H. (1980). Front-page news and real-world cues: A new look at agenda-setting by the media. *American Journal of Political Science*, *24*(1), 16–49. doi:10.2307/2110923

Fazekas, Z., & Larsen, E. G. (2016). Media content and political behavior in observational research: A critical assessment. *British Journal of Political Science*, *46*(1), 195–204. doi:10.1017/S000712341500006X

Franz, M. M., & Ridout, T. N. (2007). Does political advertising persuade? *Political Behavior*, *29*(4), 465–491. doi:10.1007/s11109-007-9032-y

Freedman, P., & Goldstein, K. (1999). Measuring media exposure and the effects of negative campaign ads. *American Journal of Political Science*, *43*(4), 1189–1208. doi:10.2307/2991823

Fuller, W. A. (1987). *Measurement error models*. New York, NY: Wiley.

Goldman, S. K., Mutz, D. C., & Dilliplane, S. (2013). All virtue is relative: A response to Prior. *Political Communication*, *30*(4), 635–653. doi:10.1080/10584609.2013.819540

Goldstein, K., & Freedman, P. (2002). Lessons learned: Campaign advertising in the 2000 elections. *Political Communication*, *19*(1), 5–28. doi:10.1080/105846002317246461

Hardin, J. W., & Carroll, R. J. (2003). Measurement error, GLMs, and notational conventions. *Stata Journal*, *3*(4), 329–341.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, *34*(1), 93–101. doi:10.2307/2092790

Hillard, D., Purpura, S., & Wilkerson, J. (2008). Computer-assisted topic classification for mixed-methods social science research. *Journal of Information Technology & Politics*, *4*(4), 31–46. doi:10.1080/19331680801975367

Hügel, R., Degenhardt, W., & Weiss, H.-J. (1989). Structural equation models for the analysis of the agenda-setting process. *European Journal of Communication*, *4*(2), 191–210. doi:10.1177/0267323189004002005

Iyengar, S. (2001). The method is the message: The current state of political communication research. *Political Communication*, *18*(2), 225–229. doi:10.1080/105846001750322998

Iyengar, S. (2011). Experimental designs for political communication research: Using new technology and online participant pools to overcome the problem of generalizability. In E. P. Bucy & R. L. Holbert (Eds.), *Sourcebook for political communication research. Methods, measures, and analytical techniques* (pp. 129–148). New York, NY: Routledge.

Iyengar, S., & Simon, A. F. (2000). New perspectives and evidence on political communication and campaign effects. *Annual Review of Psychology*, *51*(1), 149–169. doi:10.1146/annurev.psych.51.1.149

Jebril, N., de Vreese, C. H., Dalen, A. V., & Albæk, E. (2013). The effects of human interest and conflict news frames on the dynamics of political knowledge gains: Evidence from a cross-national study. *Scandinavian Political Studies*, *36*(3), 201–226. doi:10.1111/1467-9477.12003

Kepplinger, H. M., Brosius, H.-B., & Staab, J. F. (1991). Opinion formation in mediated conflicts and crises: A theory of cognitive-affective media effects. *International Journal of Public Opinion Research*, *3*(2), 132–156. doi:10.1093/ijpor/3.2.132

Kepplinger, H. M., Geiss, S., & Siebert, S. (2012). Framing scandals: Cognitive and emotional media effects. *Journal of Communication*, *62*(4), 659–681. doi:10.1111/j.1460-2466.2012.01653.x

Kleinnijenhuis, J. (1991). Newspaper complexity and the knowledge gap. *European Journal of Communication*, *6*(4), 499–522. doi:10.1177/0267323191006004006

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Thousand Oaks, CA: Sage.

LaCour, M. J., & Vavreck, L. (2014). Improving media measurement: Evidence from the field. *Political Communication*, *31*(3), 408–420. doi:10.1080/10584609.2014.921258

Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1968). *The people's choice: How the voter makes up his mind in a presidential campaign* (3rd ed.). New York, NY: Columbia University Press.

Lee, C.-J., Hornik, R., &, M. (2008). The reliability and stability of general media exposure measures. *Communication Methods and Measures*, *2*(1–2), 6–22. doi:10.1080/19312450802063024

Matthes, J. (2012). Exposure to counterattitudinal news coverage and the timing of voting decisions. *Communication Research*, *39*(2), 147–169. doi:10.1177/0093650211402322

McGuire, W. J. (1986). The myth of massive media impact: Savagings and salvagings. In G. Comstock (Ed.), *Public communication and behavior* (vol. 1, pp. 173–257). Orlando, FL: Academic Press.

Miller, A. H., Goldenberg, E. N., & Erbring, L. (1979). Type-set politics: Impact of newspapers on public confidence. *American Political Science Review*, *73*(1), 67–84. doi:10.2307/1954731

Mutz, D. C. (1992). Mass media and the depoliticization of personal experience. *American Journal of Political Science*, *36*(2), 483–508. doi:10.2307/2111487

Peter, J. (2003). Country characteristics as contingent conditions of agenda setting: The moderating influence of polarized elite opinion. *Communication Research*, *30*(6), 683–712. doi:10.1177/0093650203257844

Price, V., & Zaller, J. (1990). Evaluation of media exposure items in the 1989 NES Pilot Study. Retrieved from http://www.electionstudies.org/Library/papers/documents/nes002283.pdf

Price, V., & Zaller, J. (1993). Who gets the news? Alternative measures of news reception and their implications for research. *Public Opinion Quarterly*, *57*(2), 133–164. doi:10.1086/269363

Prior, M. (2009). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, *73*(1), 130–143. doi:10.1093/poq/nfp002

Prior, M. (2013). The challenge of measuring media exposure: Reply to Dilliplane, Goldman, and Mutz. *Political Communication*, *30*(4), 620–634. doi:10.1080/10584609.2013.819539

Romantan, A., Hornik, R., Price, V., Cappella, J., & Viswanath, K. (2008). A comparative analysis of the performance of alternative measures of exposure. *Communication Methods and Measures*, *2*(1–2), 80–99. doi:10.1080/19312450802062539

Rössler, P. (1999). The individual agenda-designing process: How interpersonal communication, egocentric networks, and mass media shape the perception of political issues by individuals. *Communication Research*, *26*(6), 666–700. doi:10.1177/009365099026006002

Scharkow, M. (2016). The accuracy of self-reported Internet use: A validation study using client log data. *Communication Methods and Measures*, *10*(1), 13–27. doi:10.1080/19312458.2015.1118446

Schemer, C. (2012). The influence of news media on stereotypic attitudes toward immigrants in a political campaign. *Journal of Communication*, *62*(5), 739–757. doi:10.1111/j.1460-2466.2012.01672.x

Schuck, A. R. T., Boomgaarden, H. G., & de Vreese, C. H. (2013). Cynics all around? The impact of election news on political cynicism in comparative perspective. *Journal of Communication*, *63*(2), 287–311. doi:10.1111/jcom.12023

Schuck, A. R. T., Vliegenthart, R., & de Vreese, C. H. (2016a). Matching theory and data: Why combining media content with survey data matters. *British Journal of Political Science*, *46*(1), 205–213. doi:10.1017/S0007123415000228

Schuck, A. R. T., Vliegenthart, R., & de Vreese, C. H. (2016b). Who's afraid of conflict? The mobilizing effect of conflict framing in campaign news. *British Journal of Political Science*, *46*(1), 177–194. doi:10.1017/s0007123413000525

Schwartz, J. (1985). The neglected problem of measurement error in categorical data. *Sociological Methods & Research*, *13*(4), 435–466. doi:10.1177/0049124185013004001

Schwarz, N., &, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, *22*(2), 127–160. doi:10.1177/109821400102200202

Shoemaker, P. J., & Reese, S. D. (1990). Exposure to what? Integrating media content and effects studies. *Journalism & Mass Communication Quarterly*, *67*(4), 649–652. doi:10.1177/107769909006700418

Slater, M. D. (2004). Operationalizing and analyzing exposure: The foundation of media effects research. *Journalism & Mass Communication Quarterly*, *81*(1), 168–183. doi:10.1177/107769900408100112

Slater, M. D. (2016). Combining content analysis and assessment of exposure through self-report, spatial, or temporal variation in media effects research. *Communication Methods and Measures*, *10*(2–3), 173–175. doi:10.1080/19312458.2016.1150969

Takens, J., Kleinnijenhuis, J., Hoof, A. V., & Atteveldt, W. V. (2015). Party leaders in the media and voting behavior: Priming rather than learning or projection. *Political Communication*, *32*(2), 249–267. doi:10.1080/10584609.2014.944319

Valkenburg, P. M., & Peter, J. (2013). Five challenges for the future of media-effects research. *International Journal of Communication*, *7*, 197–215. doi:1932–8036/2013FEA0002

van Spanje, J., & de Vreese, C. (2014). Europhile media and Eurosceptic voting: Effects of news media coverage on Eurosceptic voting in the 2009 European parliamentary elections. *Political Communication*, *31*(2), 325–354. doi:10.1080/10584609.2013.828137

Vavreck, L. (2007). The exaggerated effects of advertising on turnout: The dangers of self-reports. *Quarterly Journal of Political Science*, *2*(4), 325–343. doi:10.1561/100.00006005

Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review*, *35*(1), 112–117. doi:10.2307/2093858

Wonneberger, A., & Irazoqui, M. (2016). Explaining response errors of self-reported frequency and duration of TV exposure through individual and contextual factors. *Journalism & Mass Communication Quarterly*. Advance online publication. doi:10.1177/1077699016629372

Zaller, J. (2002). The statistical power of election studies to detect media exposure effects in political campaigns. *Electoral Studies*, *21*(2), 297–329. doi:10.1016/S0261-3794(01)00023-3