Publishing multiple journal articles from a single data set: Issues and recommendations

By: Mark A. Fine and Lawrence A. Kurdek


***Reprinted with permission. No further reproduction is authorized without written permission from the American Psychological Association. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.***

Abstract:

Researchers frequently have to decide how to publish journal articles from a large data set. When it is not possible to write a single, integrative article from such a data set, the issue of publishing multiple journal articles arises. In this article, in the context of family psychology, 2 operational criteria for determining when it is appropriate to publish multiple articles from a single data set are proposed: (1) when it is not possible to write a single integrative article that is clear, digestible, and meaningful and (2) when the multiple articles have distinct purposes. Recommendations are also made to authors who have decided to publish multiple articles from the same data set.

Keywords: publishing | publications | academic research | research publication | research data | psychology research

Article:

Because topics in the area of family psychology can be addressed from multivariate perspectives, researchers in this specialty often collect large amounts of data in any given study. The amount of data collected makes it likely that authors will consider publishing several articles from the same data set. As noted by one outgoing editor of a developmental psychology journal (Parke, 1992), issues related to multiple publication have become increasingly salient.

Authors with a large data set frequently face a dilemma. On the one hand, existing professional guidelines strongly encourage publishing single journal articles (see below). On the other hand, because few journals in the area of family psychology publish monograph-length articles, authors encounter reviewers and editors who encourage them to reduce the scope and length of their manuscripts. Thus, in writing articles based on large data sets, authors must balance the
competing pressures to combine all of their data into one article and to restrict the scope of any one given article to digestible proportions.

At the outset, we note that we endorse the view that it is generally desirable to publish one comprehensive article from a large data set. Knowledge is advanced when related research findings are presented in an integrated package. However, because diverse findings cannot always be integrated in a meaningful way in a single article that meets the page limitations of a targeted journal, we believe that there are circumstances when it is appropriate for authors to write multiple journal articles from the same data set.

The purposes of this article are (a) to help researchers determine whether it is appropriate to publish multiple journal articles from a single data set and (b) to provide guidance concerning how to prepare the multiple articles, if multiple publication is deemed appropriate. To accomplish these purposes, we review established standards for avoiding the variant of inappropriate multiple publication that is of interest in this article (i.e., “piecemeal” publication), highlight the limitations of these standards, and propose two operational criteria for determining when it is appropriate to publish several articles from a single data set. To clarify the use of these criteria, we present hypothetical cases of both appropriate and inappropriate decisions concerning multiple publication. Finally, in those instances when researchers have decided that it is appropriate to publish multiple journal articles from the same data set, we make recommendations for how to package the multiple articles.

Our focus in this article is limited in four ways. First, we address piecemeal publication and do not deal with duplicate publication. Duplicate publication occurs when an author publishes the same material in more than one journal article. With the availability of computerized systems for literature retrieval, there is no longer a need to publish the same results for different target audiences (Storandt, 1993). Because duplicate publication is relatively easy to determine and always unethical, we do not consider it further.

Second, we focus on journal articles because they are regarded in the scientific community as the primary forum for new scholarship. Books, book chapters, and other types of publications are not bound by the same standards as are journal articles because these publication outlets do not necessarily present novel information, are not necessarily peer-reviewed, are not bound by space limitations, and are not subject to guidelines from professional organizations.

Third, we focus on multiple journal articles involving empirically based articles and do not consider articles that present theoretical analyses or that critically review existing research. Issues regarding piecemeal publication are more likely to occur with empirically based articles than with non-empirically based articles.

Finally, we focus on instances when authors are responsible for the design of the study, the collection of the data, and the writing of journal articles that report their findings. We do not address issues of multiple publications related to secondary data sets. Because secondary data
sets are in the public domain (i.e., they are intended to be used by a number of researchers and are designed to spawn multiple publications), they typically contain information on a wide range of topics (e.g., the National Survey of Families and Households; Sweet, Bumpass, & Call, 1988). Because of the diverse information that these data sets contain, it is quite likely that an author could appropriately publish multiple journal articles from a secondary data set without facing a piecemeal publication dilemma.

Standards for Determining Piecemeal Publication

The American Psychological Association (APA) has three documents that provide standards regarding piecemeal publication. These are the Publication Manual of the American Psychological Association (APA, 1994), the Editor's Handbook: Operating Procedures and Policies for APA Publications (APA, 1993), and an open letter to authors who submit manuscripts to APA journals (Storandt, 1993).

The Publication Manual and the open letter to authors indicate that publication of several articles from a single data set is undesirable. If authors wish to divide the results from a single data set into multiple reports, they are instructed, at minimum, to inform the editor of this plan and to provide such information as the editor requests (e.g., copies of articles that have already been prepared from the data set). If repeated publications from a longitudinal study are warranted, the sections that describe the nature of the study and its methodology should be fully described in only the first article.

The APA's Editor's Handbook states that the key determining factor for an editor who is making a judgment regarding piecemeal publication is whether the multiple journal articles can be clearly and understandably integrated into a single article. Importantly, it is recognized that this integration may not be possible in multidisciplinary or longitudinal studies. Furthermore, according to the Editor's Handbook and the Publication Manual, data that already have been published can be recast in the context of new research questions or theoretical models.

The open letter to authors (Storandt, 1993) also states that piecemeal publications can be misleading if they are presented as representing independent data sets. When multiple articles come from a single data set, they may pose problems for researchers who review the literature or who conduct meta-analyses.

Limitations of Existing Standards Regarding Piecemeal Publication

Taken together, the existing standards indicate that the decision concerning what constitutes piecemeal publication is left to journal editors, who must make such a determination on the basis of materials provided by authors. We believe that there are three problems with this approach. First, on the basis of our collective experience in the roles of author, reviewer, and editor, issues regarding piecemeal publication most often arise when reviewers of a manuscript inform the editor of this possibility. However, because the comprehensiveness of reviewers' knowledge of
the literature(s) varies, they may not always detect possible publication problems. Consequently, the monitoring process is nonsystematic and depends on an element of chance. Although the APA's Publication Manual and the open letter to authors (Storandt, 1993) indicate that authors should inform editors when they wish to divide the results from a single data set into multiple reports, our experience suggests that this seldom takes place. As a result, editors may be unaware when piecemeal publication occurs.

Second, when editors are given relevant information from authors, no clear guidelines exist to help them determine whether a given manuscript represents piecemeal publication. Given the lack of operational criteria to guide editors, it would not be surprising to find considerable variability among editors in what they judge to be piecemeal publication. The APA's Editor's Handbook suggests that editors consult with APA staff when particularly difficult decisions are necessary. Although this may be helpful in some cases, a lack of clarity in criteria for piecemeal publication renders any judgment subjective and therefore open to interpretation.

Third, and most important, the criteria for piecemeal publication do not provide authors with the information they need to regulate their own behavior. One of the key purposes of ethical codes, standards, and manuals is to provide professionals with guidance on how to govern their activities (Keith-Spiegel & Koocher, 1985). Without clear criteria for what constitutes piecemeal publication, such guidance is not available.

Two Criteria for Assessing the Appropriateness of Multiple Journal Articles From the Same Data Set

In this section, we propose two criteria that can be used to assess the appropriateness of publishing multiple journal articles from the same data set. It should be noted that the determination of whether those criteria are met is necessarily subjective. To reduce the extent of this subjectivity, we also offer suggestions for determining whether each criterion has been met.

Criterion 1: The Multiple Articles Cannot Be Integrated Into a Single Article

The first criterion is that it is not possible to write a single article that integrates results from the entire data set. There are both practical and pedagogical aspects to this criterion. The practical aspect is that authors must not be able to write a single integrative article within the confines of the space that a particular editor or journal allows. Because there are few journals in family psychology that publish monograph-length articles, such as those published by the Society for Research on Child Development, authors typically have a limited amount of space available to them.

The pedagogical aspect of this criterion is that it must not be possible for authors to write a single, integrative article that is clear, digestible, and meaningful. An article has these qualities if it can be written in such a way that readers can readily acquire a coherent sense of the overall findings and not become excessively confused and frustrated.
Suggestions to Authors for Determining Whether Criterion 1 Is Met

Attempt to write a single, integrative article

Authors should attempt to write a single, integrative article that presents the results from the entire data set in a clear, digestible, and meaningful (i.e., effective) way.

Have colleagues assess whether the single article is clear, digestible, and meaningful

Colleagues can provide valuable input to resolve the dilemma of multiple journal articles. We suggest that authors consult with one or more colleagues who are familiar with research in family psychology about whether it is possible to write a single, integrative article from a single data set that is clear, digestible, and meaningful. If colleagues suggest that the article does not contain these qualities, they should also be asked for an opinion concerning whether it would be possible to develop such a single, integrative article with more effort. If colleagues believe that a single, effective article is possible, then authors should consider reworking the material. However, if colleagues believe that it would not be possible to write such an article, then authors should consider writing multiple journal articles or reducing the scope of the original article.

Examine journal page limitations and the lengths of typical articles

Some journals have policies that place a maximum page limit on manuscripts, and journals differ in the typical length of their published articles. If long manuscripts are not allowed or are not generally published in a targeted journal, authors may need to consider other appropriate journals, a monograph, or limiting the scope of the single article.

Consult with editor of journal

Even with feedback from colleagues, authors may still be unclear whether it is possible to write an effective article from the data. In such cases, authors might consider asking the editor of the target journal if he or she is willing to provide guidance. If authors wish to do this, we suggest that they write a single, comprehensive article and request feedback from the editor on whether the article is clear, digestible, and meaningful. The editor might also be asked whether journal page limitations can be extended when an effective article exceeds the stated page limits. Early consultation with the editor may save considerable time.

Criterion 2: Each Article Has a Distinct Purpose

The second criterion is that the multiple journal articles have distinct purposes. This criterion is met if the articles (a) address different research questions and (b) use different relevant literatures. Clearly, there is some overlap across literatures, and, in fact, there is no clear consensus on what constitutes a literature. As a working definition, we propose that a literature consists of a body of theory and knowledge on a given topic. The literature provides a context for the understanding of the purpose, importance, theoretical grounding, and possible contributions
of the article. Consequently, if the relevant literatures differ for the multiple journal articles, then the articles have distinct purposes.

Consider two examples in the area of family psychology of multiple journal articles derived from the same data set that are based on different relevant literatures. In the first example, the data set contains information on parenting cognitions, child adjustment, and family structure. One article derived from the data set focuses on parental cognitions and their relation to child adjustment and is rooted in the child socialization, parenting, child adjustment, and social cognition literatures. A second article derived from the same data set compares child adjustment in several different family structures and is embedded in the divorce, stepfamily, stress, and child development literatures.

In the second example, the data set contains information from a 30-year longitudinal data set that includes measures of psychopathology and family processes. One article examines the prevalence and stability of psychopathological symptoms over time and is embedded within the psychiatric epidemiology, psychopathology, life span development, and prevention literatures. A second article explores the relation between childhood perceptions of family processes among adults with mental disorders and is derived from the family dynamics, etiology of psychopathology, and psychotherapy literatures.

Suggestions to Authors for Determining Whether Criterion 2 Is Met

Determine whether reading multiple articles leads to a limited understanding of the results

Authors should ask themselves whether readers of any one of the multiple journal articles will acquire a limited understanding of the results of the entire study. If a reader would need to read all of the multiple articles to acquire a working understanding of the results, then the purposes of the articles are not distinct and the author should attempt to integrate the results into one article. However, if a reader could gain a thorough understanding of the results pertaining to a specific purpose in any single article, then multiple journal articles may be appropriate.

Determine the extent to which the reference lists of the multiple articles overlap

To help authors determine whether the relevant literatures are different, they should construct the reference lists that will be necessary for each of the multiple articles. If these lists overlap substantially, the relevant literatures are not different and dividing the data set into multiple articles is not appropriate. On the other hand, if the lists are distinct, the relevant literatures are different and multiple articles may be appropriate.

If there are multiple dependent variables, determine whether they are conceptually similar and empirically related

When a study contains multiple dependent variables, authors should determine whether the dependent variables are conceptually related. If they are, the criterion of distinct purposes is not
met, and separate articles based on each dependent variable are not appropriate. If the multiple
dependent variables are not conceptually related, the purposes of multiple journal articles based
on each dependent variable may be distinct. As a check on whether the dependent variables are
conceptually related, authors might examine whether they are empirically related.

Consult with colleagues

In addition to providing guidance about whether a single, integrative article is clear, digestible,
and meaningful, colleagues can also be asked whether they believe that the entire study involves
distinct purposes or whether the multiple journal articles that have already been written have
distinct purposes.

Consult with editor of journal

Authors may also wish to consult with an editor to help determine whether the criterion of
distinct purposes is met. To do this, we suggest that authors make a preliminary decision about
the distinctiveness of the multiple journal articles, provide the rationale for this decision, provide
the editor with copies of all manuscripts already generated from the data set, and ask the editor
for feedback.

In longitudinal studies, assess the importance to the scholarly community of publishing the
results from early waves of the study before the entire study is completed

Longitudinal studies provide a unique set of challenges. A relatively conservative position that
one could adopt is that researchers should wait until the completion of their longitudinal study
before they publish their results in a single, integrative article. We view this position as
untenable for three reasons: (a) longitudinal studies are often designed to address distinct
purposes, (b) it may be important to the scholarly community to obtain the results from early
waves of the study before the entire investigation is completed, and (c) because of attrition,
different phases of the study may be based on different subsets of the sample. Thus, investigators
conducting a longitudinal study should include in their evaluation (i.e., their evaluation of
whether multiple journal articles based on different waves of a study have distinct purposes) a
consideration of the importance to the scholarly community of publishing the results from early
waves before the completion of the entire study.

Coordinating the Two Criteria

If only one of the two criteria is met, we believe that it is inappropriate to publish multiple
journal articles from a single data set. There are situations when one criterion is met and the
other is not. For example, if the first criterion is not met (i.e., it is possible to write a single,
integrative article that is readable and understandable), we believe that the authors should publish
a single article, even if the article contains information relevant to distinct purposes. Thus, in the
two examples given above of multiple journal articles that are based on different literatures,
publishing multiple journal articles is justifiable only if these distinct purposes cannot be adequately integrated within the confines of a single article.

If it is not possible to write a single, integrative article (Criterion 1 is met), but the purposes addressed by the multiple articles are not distinct (Criterion 2 is not met), then we argue that it is not appropriate to publish multiple journal articles. Instances of this sort are unlikely to occur because it is typically possible to write a single, integrative article that presents the results from a single-purpose study. Nevertheless, if this scenario does occur, the recommended strategy is to continue trying to write a clear, digestible, and meaningful article. If an author eventually concludes that this is not possible, he or she should reduce the amount of data presented in the single journal article. One implication of this scenario is that there will be some data from the study that will not be published in a journal article.

An Evaluation of the Appropriateness of Publishing Multiple Journal Articles: Six Hypothetical Cases

To clarify how the two criteria can be used, we present six hypothetical cases. The first three cases illustrate instances when publishing multiple journal articles is appropriate. The last three cases illustrate instances of piecemeal publication, when publishing multiple, journal articles is not appropriate.

A Longitudinal Study

Researcher A recruited a sample of over 500 newlywed couples and planned repeated annual assessments of the couples over at least a 5-year period. Some of the measures were common to each annual assessment, whereas others were used only at specific assessments. After the first wave of data collection, she wrote an article regarding the relation between divorce history and current marital satisfaction. After three waves of data collection, she wrote an article on how couples who made the transition to parenthood differed in changes in marital satisfaction from those who did not experience the transition to parenthood. After five waves of data collection, she wrote a third article on the predictors of marital dissolution. Because of the amount of data collected and the loss of subjects over time, Professor A determined that she could not address each topic of the overall study in a single article.

The three articles written by Researcher A addressed the relation between a demographic variable (divorce history) and marital satisfaction, how changes in marital satisfaction differed between couples who had children and those who did not, and factors that predicted divorce after 5 years. In our judgment, it would have been difficult, on both practical and pedagogical grounds, to integrate these topics into one article. Thus, the first criterion was met. Was the second criterion of distinct purposes met? The particular purpose addressed in each article required different lengths of time and, therefore, different data sets to study the phenomenon of interest: The first article did not need a longitudinal design; the second article used several years of data, because any effects on marital satisfaction associated with having a child would not be
immediately apparent; and the third article used a period of 5 years, because divorce, if it occurs, typically does so after 5 or more years. Consequently, we believe that the second criterion was also met.

The Emergence of New Literature

Researcher B was interested in the relation between family process and adolescent adjustment. She conducted a survey study in which the main objective was to identify specific family processes that mediated the relation between family structure and adjustment. She wrote one article on this topic. Later, she read a newly published study that proposed that two of the family processes that she studied exert an interactive effect on adjustment. Although such an interactive effect was not of interest in the original study, in additional data analyses she found that such an effect did, indeed, occur in her sample, although in a slightly different manner than in the newly published article. She wrote a second article to address this issue.

At the time that the first article was written, Researcher B was not interested in the possibility of an interactive effect between the two family processes of interest, and there was no available literature that indicated that this effect was a likely possibility. However, after finding such an effect in her data set, she chose to publish this finding to replicate and to extend the results from the published article. We believe that her decision to publish the second article was appropriate, because the second paper addressed a purpose that was distinct from that of the first article (Criterion 2). Of course, had she known that a previous study had found an interactive effect before she wrote the first article, she should have packaged her results into one article, if it were possible to write a single article that was clear, digestible, and meaningful (Criterion 1).

The Use of Different Subsamples in the Data Set

Researcher C was interested in the relation between family structure and child adjustment. She administered a survey to junior high school students. She discovered that there were large numbers of students in the two-biological parent, single divorced mother, and stepfather families, but very small numbers of students in the single divorced father and stepmother families. As part of the survey, students in stepfamilies answered survey items regarding the parenting behaviors of their stepparents as well as of their parents. She wrote two papers. The first paper examined differences in adjustment in children from two-biological parent, single divorced mother, and stepfather families (the groups with the largest sample sizes). To facilitate the interpretation of family structure effects, she decided to exclude the rather small number of students in the single divorced father and stepmother families. In the second paper, she explored whether children's adjustment was related to how they rated the parenting behaviors of their parents and stepparents.

Researcher C wrote two articles that could not easily be integrated into a single journal article (Criterion 1) and that had distinct purposes (Criterion 2). The first article was designed to compare children's well-being in several family structures and was grounded in the child
development and stress literatures. The second article was more focused than the first and, among only those children in stepfamilies, examined the link between children's perceptions of (step)parental behaviors and their adjustment. This article was rooted in the stepfamily, parental socialization, and family process literatures. In addition, the second article used data (ratings of stepparents) that could not be used in the first one, because most children included in the first article did not have stepparents. As a result of these considerations, we believe that the publication of multiple articles from this data set was appropriate.

As discussed further below, the author could label the two articles with a Roman numeral “I” and a Roman numeral “II” to acknowledge that they were generated from the same data set and submit them to the same journal, with appropriate cross-references in each article to the other. If feasible, the two articles could be submitted to the journal simultaneously.

Conceptually Similar and Empirically Related Dependent Variables in a Longitudinal Study

Researcher D was interested in changes in marital quality and recruited a sample of newlywed couples and annually collected data from them over a 5-year period. The major question of interest was the extent to which personality traits at the time of marriage predict marital satisfaction and marital commitment at 5 years postmarriage. The researcher wrote two papers. One focused on the prediction of marital satisfaction, and the other focused on the prediction of marital commitment. The design of the study was the same in each paper, with the exception of the change in the outcome measure.

Researcher D wrote two papers from the same data set that focused on conceptually similar dependent variables: marital satisfaction and marital commitment. We believe that both marital outcomes could have been included in a clear, digestible, and meaningful single article (Criterion 1). On practical grounds, the inclusion of both dependent variables would not take much additional space beyond that necessary to include only one dependent variable. On pedagogical grounds, the conceptual similarity between marital satisfaction and marital commitment would make it relatively easy to clearly present and integrate the results from these two variables. Furthermore, because of the conceptual similarities between marital satisfaction and commitment and the overlap in their respective literatures, the purposes of the two papers were not distinct (Criterion 2). In addition, it is quite likely that the marital satisfaction and marital commitment variables would be highly intercorrelated, which would provide further justification for the researcher including both variables in a single article. Thus, we believe that multiple articles were inappropriate and constituted piecemeal publication.

Different Subsamples but Similar Literature and Research Questions

Researcher E was interested in the extent to which support for the marriage from family and friends affects ratings of relationship quality. He collected data from a large convenience sample. Although not specifically planned, the sample contained about 60% of persons in their first marriages and 40% of persons in remarriages. The researcher prepared two papers that examined
the question of interest. In the first paper, he used data from the first-married subjects, whereas in
the second paper he used data from the remarried subjects. The variables assessed and reported
in the two papers were the same.

Researcher E wrote two papers that addressed the same basic research question: the relation
between support from family and friends and relationship quality. Data from respondents in first
marriages and remarriages on both practical and pedagogical grounds could easily have been
included in the same article (Criterion 1). Furthermore, because the relevant literatures were the
same (social support and marital quality) the criterion of distinct purposes was not met. Without
a convincing conceptual reason to separate the results from participants in first-marriages and
those in remarriages, we believe that these two articles provide an example of piecemeal
publication.

Different Subsamples, but Similar Research Questions, Measures, and Theoretical Frameworks

Researcher F was interested in rates of divorce in first-marriages and remarriages in different
countries. She obtained data from an international health organization that contained information
on a number of different countries and wrote two papers. The first addressed the divorce rate in
these two kinds of couples in Canada, and the second dealt with the same issue in England. The
same theoretical rationale and measures were used in both papers, although the results led to
different conclusions.

Because the two articles that Researcher F wrote involved similar research questions, measures,
and theoretical frameworks, the criterion of distinct purposes was not met. Although the results
were different in the two countries, the findings should have been published as a single,
integrative article (Criterion 1). As in the two previous cases, on both practical and pedagogical
grounds, a clear, digestible, and meaningful single article could have been written. The results
from the different countries are informative in a comparative sense and are likely to shed light on
the issue of divorce rates that neither of the individual articles could. In addition, because data
were available from multiple countries, we believe that the investigator should have determined a
priori (on the basis of her conceptual framework) which countries she wished to examine in the
single article.

Recommendations to Authors Who Have Concluded That Multiple Journal Articles Are Appropriate

Once an author has decided that it is appropriate to publish multiple journal articles, we
recommend that the following practices be observed.

Cross-Reference the Multiple Articles

As the Publication Manual and the Editor's Handbook (p. 1.6) indicate and as stated in the open
letter to authors (Storandt, 1993), multiple journal articles from the same data set should
acknowledge each other. By doing so, the author informs the scholarly community that the articles were generated from the same data set. In addition, by acknowledging the previous work(s), authors can make descriptions of the methodology of the study briefer in subsequent articles and save journal space.

Stipulate in a Cover Letter to the Editor That the Manuscript Submitted Is One of Several From the Same Data Set

We suggest that authors indicate in their cover letters to journal editors that the manuscript submitted is one of several from the same data set. This action serves to alert the editor to the existence of multiple journal articles and satisfies the guidelines in the Publication Manual (1994, p. 297) and the Editor's Handbook (pp. 1.8–1.9).

Develop a Primary Article and One or More Secondary Articles

Concerning several articles that are based on the same data set, we recommend that one article be considered as the primary article and that the results from specialized analyses of some portion of the data be presented in secondary articles. Each of the secondary articles should cite the primary article as the source for detail about the methodology of the study. A commonly used convention is to designate multiple articles (in the same journal) from the same data set with consecutive Roman numerals. For example, the title of the primary article might include a “I,” the second a “II,” and so on.

Defend the Distinctiveness of the Current Article

In the article itself, authors should specify how the current manuscript is distinct from the previous articles from the same data set.

Submit Previous Papers With the Current Manuscript

As suggested by APA's Publication Manual (1994, p. 297), the Editor's Handbook (p. 1.7), and the open letter to authors (Storandt, 1993), when submitting manuscripts to editors, authors should submit manuscripts, preprints, and reprints of other papers from the same data set. Because our impression is that authors seldom follow this practice, organizations such as the APA that publish scholarly works and editors need to educate the scholarly community. We support the APA's practice of including this guideline in the Instructions for Authors section of APA journals.

Conclusion

Our aim in this article was to bring some clarity to the issue of multiple publication in the context of empirically based journal articles from author-designed data sets. We chose to focus on this context because we believe that multiple publication dilemmas occur most frequently and are most complex in these situations. However, we suspect that the issues raised in this article are
also applicable to non–empirically based journal articles and to secondary data sets. We hope that others in the family psychology scholarly community will use our ideas as a catalyst to evaluate the usefulness and generalizability of our proposed criteria and recommendations.

References


