

# Quality of Reporting of Diagnostic Accuracy Studies: No Change Since STARD Statement Publication—Before-and-after Study<sup>1</sup>

Nancy L. Wilczynski, MSc, PhD

**Purpose:**

To determine the quality of reporting of diagnostic accuracy studies before and after the Standards for Reporting of Diagnostic Accuracy (STARD) statement publication and to determine whether there is a difference in the quality of reporting by comparing STARD (endorsing) and non-STARD (nonendorsing) journals.

**Materials and Methods:**

Diagnostic accuracy studies were identified by hand searching six STARD and six non-STARD journals for 2001, 2002, 2004, and 2005. Diagnostic accuracy studies ( $n = 240$ ) were assessed by using a checklist of 13 of 25 STARD items. The change in the mean total score on the modified STARD checklist was evaluated with analysis of covariance. The change in proportion of times that each individual STARD item was reported before and after STARD statement publication was evaluated ( $\chi^2$  tests for linear trend).

**Results:**

With mean total score as dependent factor, analysis of covariance showed that the interaction between the two independent factors (STARD or non-STARD journal and year of publication) was not significant ( $F = 0.664$ ,  $df = 3$ , partial  $\eta^2 = 0.009$ ,  $P = .58$ ). Additionally, the frequency with which individual items on the STARD checklist were reported before and after STARD statement publication has remained relatively constant, with little difference between STARD and non-STARD journals.

**Conclusion:**

After publication of the STARD statement in 2003, the quality of reporting of diagnostic accuracy studies remained similar to pre-STARD statement publication levels, and there was no meaningful difference (ie, one additional item on the checklist of 13 of 25 STARD items being reported) in the quality of reporting between those journals that published the STARD statement and those that did not.

© RSNA, 2008

<sup>1</sup> From the Health Information Research Unit, McMaster University, 1200 Main St West, HSC-3H7, Hamilton, ON, Canada L8N 3Z5. Received November 28, 2007; revision requested February 1, 2008; revision received February 8; accepted April 1; final version accepted April 10. Supported by the U.S. National Library of Medicine and Canadian Institutes of Health. Address correspondence to the author (e-mail: wilczyn@mcmaster.ca).

© RSNA, 2008

**B**efore incorporating evidence into practice or including the results of a study in a systematic review, the clinician or researcher must assess the validity of the study results, as flaws in research design can lead to biased results. Previous research has shown that diagnostic test accuracy may be overestimated in studies having methodological shortcomings (1–3). To guard against the determination of decisions on the basis of biased information, the clinician or researcher must critically appraise the evidence. This appraisal can be difficult because many clinicians and researchers lack critical appraisal skills (4) and, for those who do have adequate skills, the substandard reporting of many published research reports hampers the critical appraisal exercise. A survey of diagnostic accuracy studies published in four major medical journals between 1978 and 1993 showed that the methodological quality of the studies was mediocre at best (5). The authors of the survey, in addition, reported that the quality assessments were hampered because many reports lacked key information about important design elements (5).

Awareness of the need to improve the quality of reporting of diagnostic accuracy studies was increased with the publication of the Standards for Report-

ing of Diagnostic Accuracy (STARD) statement (6,7). The STARD statement provides guidelines for improving the quality of reporting of studies of diagnostic accuracy and draws on 33 previously published checklists (8). It is hypothesized that complete and informative reporting will lead to better decisions in health care.

Similar initiatives have been launched previously for the reporting of therapy articles (Consolidated Standards of Reporting Trials [CONSORT]) (9) and systematic reviews (Quality of Reporting of Meta-analyses, known as QUOROM [10]). In addition, studies (11,12) in which the effect of these initiatives was evaluated have been published. For example, Moher et al (11) found that the use of the CONSORT statement was associated with improvements in the quality of reports of randomized controlled trials.

To date, researchers have published several studies in which the quality of reporting of diagnostic accuracy studies by using the STARD checklist was summarized (13–15). In only two studies (16,17) did the researchers investigate whether the quality of reporting had improved since the STARD statement was published. In one study (16), researchers found a slight improvement in the quality of reporting for only a few items on the STARD checklist, whereas in the other study (17), researchers found no improvement. Both studies conducted their post-STARD statement publication evaluation in 2004, however, which may have been too soon to determine the full impact of the STARD statement because the first journals to publish the statement did so in 2003.

In view of the mounting interest in evidence-based care, including diagnostic test information and the rise of systematic reviews of diagnostic accuracy studies, evaluation of the quality of reporting in diagnostic studies and the potential impact of the STARD initiative is important. The purpose of this study was to determine the quality of reporting of diagnostic accuracy studies before and after STARD statement publication and to determine whether there is a difference in the quality of reporting by comparing STARD (endorsing) and non-STARD (nonendorsing) journals.

## Materials and Methods

This research was funded by the U.S. National Library of Medicine and Canadian Institutes of Health. The funding source had no involvement in the collection, analysis, or interpretation of data or in the writing of the report.

The Figure outlines the study design. Briefly, diagnostic accuracy studies were identified by hand searching six journals that published the STARD statement in 2003 and six journals that did not publish the STARD statement. The 12 journals that were hand searched were identified by using data from the Clinical Hedges Study (18,19). Of the 170 journals included in the Clinical Hedges database, six published the STARD statement in 2003 (herein referred to as the STARD journals). These six journals were *American Journal of Roentgenology* and *Radiology* (two journals about diagnostic radiology); *Annals of Internal Medicine* (one journal about internal medicine); and *BMJ*, *JAMA*, and *Lancet* (three journals about general medicine). Five journals that had a similar content area did not publish the STARD statement. These five journals were *Pediatric Radiology* (one journal about diagnostic radiology); *Archives of Internal Medicine* (one journal about internal medicine); and *American Journal of Medicine*, *British Journal of General Practice*, and *New England Journal of Medicine* (three journals about general medicine).

To locate an additional comparator journal in the diagnostic area, as one was

## Advances in Knowledge

- After publication of the STARD statement in 2003, the quality of reporting of diagnostic accuracy studies remained similar to levels before STARD statement publication, and there was no meaningful difference (ie, one additional item on the checklist of 13 of 25 STARD items being reported) in the quality of reporting between those journals that published the STARD statement and those that did not in the first 2 years after publication.
- Quality of reporting of diagnostic accuracy is substandard, with a great deal of room for improvement.

### Published online

10.1148/radiol.2483072067

**Radiology 2008**; 248:817–823

### Abbreviations:

CONSORT = Consolidated Standards of Reporting Trials  
STARD = Standards for Reporting of Diagnostic Accuracy

### Author contribution:

Guarantor of integrity of entire study; N.L.W.

### Funding:

This research was supported by the National Library of Medicine.

Author stated no financial relationship to disclose.

See also the editorial by Bossuyt in this issue.

not available in the Clinical Hedges database, we used a method similar to that of Moher et al (11) when choosing a journal subset for evaluating the effect of the CONSORT statement for reporting clinical trials. Criteria for choosing the sixth comparator journal were as follows: (a) The journal was identified by the journal subject category of "Radiology, Nuclear Medicine and Medical Imaging" and was chosen on the Institute for Scientific Information Journal Citation Reports Web site at <http://apps.isiknowledge.com/>. (b) The journals in this subject category were sorted in descending order according to journal impact factor. (c) The first journal that was available through the Health Sciences Library at McMaster University, Hamilton, Ontario, Canada, that published at least one diagnostic accuracy study (after the contents of two 2001 issues of the journal were reviewed) and was indexed in both MEDLINE and EMBASE was identified. *European Radiology* was chosen by using these criteria, and this journal completes the set of the six comparator journals herein referred to as non-STARJ journals.

Four years of publication were studied independently: The years 2001 and 2002 were studied to obtain a pre-STARJ statement publication assessment; 2004 and 2005 were studied to obtain a post-STARJ statement publication assessment. Diagnostic accuracy studies were defined as those in which the outcomes from one or more tests considered for evaluation were compared with outcomes from the reference standard, both measured in the same study population (7). Three trained research assistants, who all had master's degree-level training in critical appraisal, participated in three training sessions held during 1 month. In the training sessions, the research assistants discussed and honed application of the definition for diagnostic accuracy studies. Crude agreement for identifying those studies was also calculated. The research assistants independently reviewed all issues of the 12 journals for the 4 years of publication to identify all diagnostic accuracy studies.

Identified diagnostic accuracy stud-

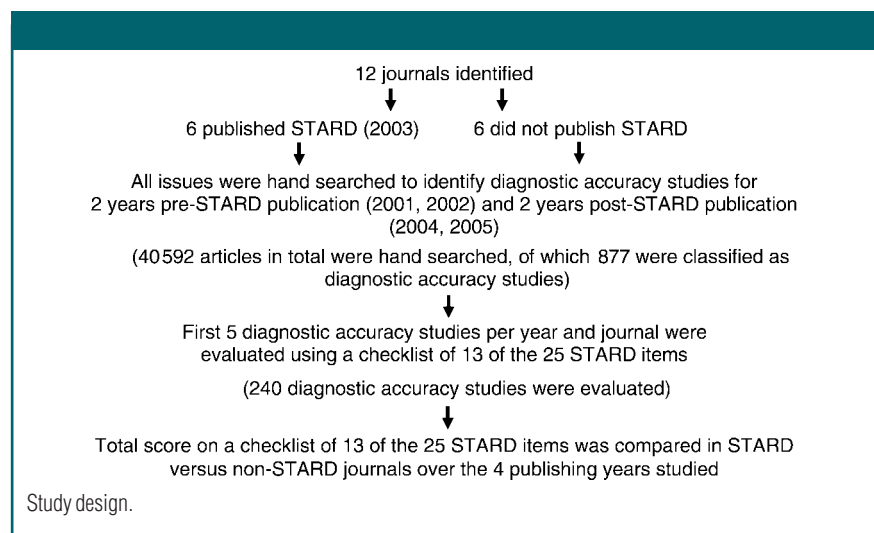
ies were evaluated by using the STARJ checklist by assigning a yes or no response to each item, depending on whether the authors had reported it. Twenty-five items make up the STARJ checklist. When the papers considered in this study were assessed, items 13 ("Methods for calculating test reproducibility, if done"), 23 ("Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done"), and 24 ("Estimates of test reproducibility, if done") were removed from the STARJ checklist because all have the qualifier "if done." Thus, if these items were not reported in the diagnostic accuracy papers evaluated, it would be impossible to determine whether this lack of reporting was because the item was not done or because it was not reported.

In addition, only those STARJ items that have been empirically shown to have a potentially biasing effect on the results of diagnostic accuracy studies and those items that appear to account for variation between studies were evaluated. On the basis of the findings in three studies (1–3), the following features, with corresponding STARJ checklist item number, appear to have had a biasing effect or account for variation between studies and should therefore be reported: (a) a description of the distribution of severity of disease of the study population (item 18), (b) a description of the reference standard and test (items 7–10),

(c) whether blinding to results of the interpretation of the test and the reference standard was done (item 11), (d) a description of the study population (items 3–5 and 14–16), and (e) a description of the data collection process (item 6). Thus, only these 13 items were assessed in this study. The checklist of 13 of the 25 STARJ items we used is shown in Table 1 because only these items have been empirically shown to have had a potentially biasing effect on the results of diagnostic accuracy studies or to account for variation among studies.

Items 8–11 on the checklist concern the index tests, as well as the reference standard. These four items were split into two questions, one for the index tests and one for the reference standard. A total STARJ score for each article was calculated by summing the number of reported items (0–13 points available). Unit weights were applied for each of the items except for items 8–11 that were split in two and that were given a weight of 0.5 for each of the index tests and the reference standard.

Four trained research assistants who all had master's degree-level training in critical appraisal (three of these research assistants were involved in identifying diagnostic accuracy studies) participated in three training sessions held to review in detail the application of the STARJ checklist to 21 articles. These research assistants independently evaluated the di-



agnostic accuracy articles (assigned randomly) in duplicate (ie, two research assistants per article). Disagreements on checklist items were resolved in consensus involving a third trained research assistant who was one of the four not involved in the duplicate assessment of the article in question. Consensus evaluations were used in the analysis. Crude agree-

ment levels were calculated for each of the 13 items after a series of 42 consecutive articles was assessed. Once crude agreement for all 13 items had reached more than 95% for at least two of the four readers, these two readers independently evaluated the remaining articles. Thus, 118 articles were assessed in duplicate (2124 individual items assessed with 55

disagreements), with the consensus evaluations used in the analysis, and 122 articles were assessed by one or the other of two readers after they attained more than 95% crude agreement.

By using software (SPSS, version 13; SPSS, Chicago, Ill), the total score on the STARD checklist was treated as the dependent variable in an analysis of

Table 1

### Reporting of Items on a Checklist of STARD Items for Evaluation of Studies of Diagnostic Accuracy according to Year of Publication

Section and Item No.	Item Description	No. of Articles per Year*			
		2001	2002	2004	2005
<b>Methods</b>					
3	The study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	28 (47)	26 (43)	35 (58)	25 (42)
4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	55 (92)	52 (87)	55 (92)	57 (95)
5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.	28 (47)	32 (53)	33 (55)	32 (53)
6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	34 (57)	40 (67)	43 (72)	44 (73)
7	The reference standard was performed and its rationale.	43 (72)	39 (65)	44 (73)	47 (78)
8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references				
	a) For index tests.	51 (85)	50 (83)	45 (75)	50 (83)
	b) For reference standard.	36 (60)	32 (53)	33 (55)	28 (47)
9	Definition of and rationale for the units, cutoffs, and/or categories of the results				
	a) Of the index tests.	38 (63)	41 (68)	42 (70)	43 (72)
	b) Of the reference standard.	33 (55)	32 (53)	32 (53)	35 (58)
10	The number, training, and expertise of the persons executing and reading				
	a) The index tests.	22 (34)	16 (27)	18 (30)	23 (38)
	b) The reference standard.	11 (18)	12 (20)	10 (17)	14 (23)
11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers				
	a) For index tests.	32 (53)	39 (65)	38 (63)	36 (60)
	b) For reference standard.	25 (42)	32 (53)	29 (48)	28 (47)
<b>Results</b>					
14	When study was done, including beginning and ending dates of recruitment.	40 (67)	41 (68)	43 (72)	45 (75)
15	Clinical and demographic characteristics of the study population (eg, age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).	52 (87)	53 (88)	57 (95)	55 (92)
16	The number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test.	26 (43)	35 (58)	29 (48)	35 (58)
	a) Flow diagram (not counted in total score).	4 (7)	9 (15)	12 (20)	14 (23)
18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	41 (68)	37 (62)	45 (75)	36 (60)

Note.—The checklist for this study included 13 of 25 items in the STARD checklist.

\* Numbers in parentheses are percentages, which were calculated on the basis of 60 articles per year of publication.

covariance with two independent factors, journal type (STARD vs non-STARD) and year of publication (2001, 2002, 2004, and 2005), with one covariate, content area of the journal (ie, diagnostic radiology vs other [general and internal medicine]). A critical level of  $P < .05$  was used to define significance. Effect sizes, as an estimate of the proportion of variance in the dependent variable (ie, total score on the STARD checklist) explained by the independent variables (ie, journal type and year of publication), are expressed as partial  $\eta^2$ . By using the software, a similar analysis of covariance was also conducted, with year of publication as a binary variable (pre-STARD [years 2001 and 2002] and post-STARD [years 2004 and 2005] statement publication) to increase the statistical power to detect a difference in total scores.

Norman and Streiner (20) provide two methods for calculating sample size when conducting an analysis of covariance that is based on the work of Kleinbaum and colleagues (21). The second method is most relevant for determining the sample size in this study and involves use of the formula for a  $t$  test and calculation of the sample size on the basis of the comparison of most interest and, in this study, that comparison is the difference in total score between STARD and non-STARD journals. By using research tool software (Arcus QuickStat Biomedical; Research Solutions, Cambridge, England), a mean difference score of 2 (a difference score of 2 would be meaningful because all 13 items on the selected checklist of STARD criteria have been empirically shown to have a biasing effect on diagnostic accuracy studies or account for the variation between studies) and a standard deviation of 5.1 (the standard deviation of the total score on a 40-item CONSORT scale in 1994 in CONSORT adopters journals [11]), 102 articles per group ( $\alpha = .05$ ,  $\beta = .20$ ), are required. On the basis of this sample size calculation, 102 articles are required in each of the STARD and non-STARD journal sets.

To provide equal samples within each journal title across 4 years of sam-

Table 2

## Number of Diagnostic Accuracy Studies according to Journal Title and Year

Subset and Journal Title	No. of Diagnostic Accuracy Studies per Year			
	2001	2002	2004	2005
<b>STARD journals</b>				
<i>AJR American Journal of Roentgenology</i>	63	75	61	48
<i>Radiology</i>	24	85	83	69
<i>Annals of Internal Medicine</i>	3	3	8	5
<i>BMJ</i>	5	4	3	4
<i>Journal of the American Medical Association</i>	6	5	3	12
<i>Lancet</i>	13	5	5	4
Total	114	177	163	142
<b>Non-STARD journals</b>				
<i>European Radiology</i>	33	31	60	46
<i>Pediatric Radiology</i>	9	12	4	9
<i>Archives of Internal Medicine</i>	5	7	3	4
<i>American Journal of Medicine</i>	2	11	12	4
<i>British Journal of General Practice</i>	4	4	1	3
<i>New England Journal of Medicine</i>	5	3	4	5
Total	58	68	84	71

pling, five diagnostic accuracy articles per journal per year of publication are required, leading to a total of 240 articles. The first five diagnostic accuracy articles in each journal for each year of publication were chosen for evaluation with the checklist we developed that was based on STARD criteria. When this was not possible (ie, when fewer than five diagnostic accuracy articles were published in a given journal within a given year), additional diagnostic accuracy articles were included from the other journals in that year and the journal subset (ie, STARD journals or non-STARD journals).

By using the research tool software, the proportion of times that each of the 13 STARD items were reported across the 4 years of publication was also analyzed by using the Armitage  $2 \times 4$  (4 years of publication)  $\chi^2$  test for linear trend. A critical level of  $P < .05$  was used to define significance.

## Results

Of the 40 592 articles reviewed, 877 were classified as diagnostic accuracy studies. Crude agreement of 96% in classifying 428 articles as diagnostic accuracy studies was attained by the three research assistants. Table 2 shows the number of diagnostic accuracy studies

for each of the 12 journals that were hand searched according to the year of statement publication. Most of the studies on diagnostic accuracy were published in three of the diagnostic radiology journals (*American Journal of Roentgenology*, *Radiology*, and *European Radiology*).

In total, 240 diagnostic accuracy studies were evaluated by using a checklist of 13 of the 25 STARD items; there were 30 articles per year (2001, 2002, 2004, and 2005) per journal subset (STARD and non-STARD journals). The mean total scores on the checklist of 13 of 25 STARD items, according to journal subset and year of publication, are shown in Table 3.

By using the mean total score as the dependent factor, an analysis of covariance showed that the interaction between the two independent factors (STARD or non-STARD journal and year of publication [a categorical factor with four levels]) was not significant ( $F = 0.664$ ,  $df = 3$ ,  $P = .58$ ). Partial  $\eta^2$  was 0.009, which means that 0.9% of the variability in the dependent variable can be explained or accounted for by the independent variables. This nonsignificant result means that the change in the mean total score in STARD journals was not significantly different from the change in the mean total score in non-

STARD journals during the 4 years of publication. The main effects for journal type (STARD vs non-STARD) ( $F = 0.897$ ,  $df = 1$ , partial  $\eta^2 = 0.004$ ,  $P = .35$ ) and journal year of publication ( $F = 0.789$ ,  $df = 3$ , partial  $\eta^2 = 0.010$ ,  $P = .50$ ) were also not significant. All effects (interaction and main effects) were also nonsignificant in the analysis where year of publication was considered as a binary variable.

Because there was no statistical difference between STARD and non-STARD journals, Table 1 shows the reporting of individual items according to year of publication. By using the proportion of times that each item was reported across the 4 years of publication, the Armitage  $\chi^2$  test for linear trend showed no significant overall differences or linear trends for any item except for the inclusion of a flow diagram (affiliated with item 16 on the STARD checklist), which is a diagram showing the exact number of patients at each stage of the study and the number of subjects who failed to undergo either the index test or the reference standard ( $\chi^2$  for linear trend,  $P = .01$ ).

## Discussion

After publication of the STARD statement, the quality of reporting remained similar to pre-STARD statement publication levels. In addition, there is no meaningful difference (ie, one additional item on the checklist of 13 of the 25 STARD items being reported) in the quality of reporting between those journals that published the STARD statement and those that did not. One possible explanation for the lack of improvement after STARD statement publication is that evaluation was con-

ducted too soon for an improvement to be evident. Evaluation of the impact of the STARD statement 2 years after publication seemed reasonable, however, because the CONSORT statement showed an improvement at 2 years after publication (11). Perhaps, as some authors suggest, the design aspects of therapeutic trials were better known and easier to apply when the CONSORT statement was published, and this same level of knowledge and ease of application was not apparent for diagnostic accuracy study principles when the STARD statement was published (22). Evaluation of the impact of the STARD statement several years after publication may be needed, particularly since the time from manuscript submission to article publication can be on the order of 12–15 months and, in some cases, longer.

Another possible explanation for the lack of improvement after STARD statement publication is that, in the journals that published the STARD statement, the editors and the peer reviewers are not enforcing the statement's guidelines. When reviewing the instructions to authors for each of the STARD journals online, it is difficult to determine the extent of enforcement of the guidelines, and it is impossible to determine whether the enforcement has changed during the 2 years of publication after the STARD statement publication.

A further possible explanation for the lack of a difference before and after STARD statement publication and the lack of a difference between STARD and non-STARD journals is that, overall, there was a general awareness of diagnostic accuracy design principles. This awareness could be the case be-

cause the STARD checklist drew on 33 previously published checklists, making reporting guidelines widely available to all authors before and after STARD statement publication. However, we found low rates of adherence to the STARD checklist items, so awareness would then not equate to adherence.

Smidt et al (16) also did not find a difference in reporting between STARD and non-STARD journals. However, they did find a slight improvement in the reporting of seven STARD items, four of which were evaluated in this study. One additional difference between our study and the study of Smidt et al (16) is that there are some marked differences in the frequency with which various items are reported. For example, Smidt et al found that item 9a, "Definition of and rationale for the units, cutoffs and/or categories of the results of the index tests" was reported 83%–94% of the time, whereas the corresponding figures from our study were 63%–71%. In conducting this study, we found that many of the STARD checklist items were open to interpretation and that many of them were multifaceted. Thus, the evaluation process for each item could vary between studies, leading to differences in the apparent frequency of item reporting.

This study had several limitations. First, it is possible that those journals classified as non-STARD journals did employ the STARD criteria even though they did not publish the STARD statement. This would make it difficult to find a difference in the quality of reporting between STARD and non-STARD journals. Second, including some journals that are nondiagnostic in focus in this study sample may have made it more difficult to discern whether the STARD statement has had an impact on the quality of reporting of diagnostic accuracy studies. Third, in this study, diagnostic accuracy studies were assessed by using only 13 of the 25 STARD criteria. This may have caused underestimation of the impact of the STARD statement publication and may have caused difficulty in the comparison of the results of this study with results of other studies in which the quality of reporting

**Table 3**

### Mean Total STARD Score according to Journal Subset and Year of Publication

Subset	Mean Total STARD Score per Year*			
	2001	2002	2004	2005
STARD journals	8.12 (2.49)	7.82 (3.09)	8.72 (2.63)	8.73 (2.29)
Non-STARD journals	7.65 (2.42)	8.28 (2.08)	8.17 (2.28)	8.10 (1.55)

\* Score was based on a checklist of 13 of the 25 STARD items and a maximum total score of 13. Numbers in parentheses are the standard deviations.

was evaluated. In addition, some may not agree with the decision to assess only those items that were empirically shown to have a biasing effect or to account for the variation between studies.

Overall, our study results show that the quality of reporting on many of the items that have been shown to have a biasing effect on the results of diagnostic accuracy studies or appear to account for the variation between studies is substandard, making it difficult for clinicians and researchers to judge study validity. Authors, editors, and peer reviewers are encouraged to adhere to and enforce STARD statement guidelines because there is clearly room for improvement in the reporting of diagnostic accuracy studies. Research is needed to determine the factors that influence journal editor and author uptake of reporting guidelines.

**Acknowledgments:** Brian Haynes, MD, PhD, Stephen Walter, PhD, and Ruta Valaitis, PhD, were involved in the development of the research protocol. Cynthia Lokker, PhD, Monika Kastner, MSc, and Leslie Baier, BA, assisted with data collection.

## References

- Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-1066.
- Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189-202.
- Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469-476.
- Putnam W, Twohig PL, Burge FI, Jackson LA, Cox JL. A qualitative study of evidence in primary care: what the practitioners are saying. *CMAJ* 2002;166:1525-1530.
- Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research: getting better but still not good. *JAMA* 1995;274:645-651.
- Bossuyt PM, Reitsma JB; Standards for Reporting of Diagnostic Accuracy. The STARD initiative [letter]. *Lancet* 2003;361:71.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003;49:7-18.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. Toward complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Fam Pract* 2004;21:4-10.
- Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-639.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement—quality of reporting of meta-analyses. *Lancet* 1999;354:1896-1900.
- Moher D, Jones A, Lepage L; CONSORT Group (Consolidated Standards for Reporting of Trials). Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;285:1992-1995.
- Devereaux PJ, Mannes BJ, Ghali WA, Quan H, Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist. *Control Clin Trials* 2002;23:380-388.
- Smidt N, Rutjes AW, van der Windt AW, et al. Quality of reporting of diagnostic accuracy studies. *Radiology* 2005;235:347-353.
- Selman TJ, Khan KS, Mann CH. An evidence-based approach to test accuracy studies in gynecologic oncology: the "STARD" checklist. *Gynecol Oncol* 2005;96:575-578.
- Siddiqui MA, Azuara-Blanco A, Burr J. The quality of reporting of diagnostic accuracy studies published in ophthalmic journals. *Br J Ophthalmol* 2005;89:261-265.
- Smidt N, Rutjes AW, van der Windt DA, et al. The quality of diagnostic accuracy studies since the STARD statement: has it improved? *Neurology* 2006;67:792-797.
- Coppus SF, van der Veen F, Bossuyt PM, Mol BW. Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. *Fertil Steril* 2006;86:1321-1329.
- Haynes RB, Wilczynski NL. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. *BMJ* 2004;328:1040-1042.
- Wilczynski NL, Haynes RB; Hedges Team. EMBASE search strategies for identifying methodologically sound diagnostic studies for use by clinicians and researchers. *BMC Med* 2005;3:7.
- Norman GR, Streiner DL. *Biostatistics: the bare essentials*. 2nd ed. Hamilton, Ontario, Canada: BC Decker, 2000.
- Kleinbaum DG, Kupper LL, Muller KE. *Applied regression analysis and other multivariable methods*. 2nd ed. Boston, Mass: PWS-Kent, 1988.
- Reeves BC. Evidence about evidence. *Br J Ophthalmol* 2005;89:253-254.

# Radiology 2008

## This is your reprint order form or pro forma invoice

(Please keep a copy of this document for your records.)

Reprint order forms and purchase orders or prepayments must be received 72 hours after receipt of form either by mail or by fax at 410-820-9765. It is the policy of Cadmus Reprints to issue one invoice per order.

**Please print clearly.**

Author Name \_\_\_\_\_  
Title of Article \_\_\_\_\_  
Issue of Journal \_\_\_\_\_ Reprint # \_\_\_\_\_ Publication Date \_\_\_\_\_  
Number of Pages \_\_\_\_\_ KB # \_\_\_\_\_ Symbol Radiology  
Color in Article? Yes / No (Please Circle)

**Please include the journal name and reprint number or manuscript number on your purchase order or other correspondence.**

### Order and Shipping Information

#### Reprint Costs (Please see page 2 of 2 for reprint costs/fees.)

\_\_\_\_\_ Number of reprints ordered \$ \_\_\_\_\_  
\_\_\_\_\_ Number of color reprints ordered \$ \_\_\_\_\_  
\_\_\_\_\_ Number of covers ordered \$ \_\_\_\_\_  
**Subtotal** \$ \_\_\_\_\_  
Taxes \$ \_\_\_\_\_

(Add appropriate sales tax for Virginia, Maryland, Pennsylvania, and the District of Columbia or Canadian GST to the reprints if your order is to be shipped to these locations.)

First address included, add \$32 for  
each additional shipping address \$ \_\_\_\_\_

**TOTAL** \$ \_\_\_\_\_

#### Shipping Address (cannot ship to a P.O. Box) Please Print Clearly

Name \_\_\_\_\_  
Institution \_\_\_\_\_  
Street \_\_\_\_\_  
City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_  
Country \_\_\_\_\_  
Quantity \_\_\_\_\_ Fax \_\_\_\_\_  
Phone: Day \_\_\_\_\_ Evening \_\_\_\_\_  
E-mail Address \_\_\_\_\_

#### Additional Shipping Address\* (cannot ship to a P.O. Box)

Name \_\_\_\_\_  
Institution \_\_\_\_\_  
Street \_\_\_\_\_  
City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_  
Country \_\_\_\_\_  
Quantity \_\_\_\_\_ Fax \_\_\_\_\_  
Phone: Day \_\_\_\_\_ Evening \_\_\_\_\_  
E-mail Address \_\_\_\_\_

\* Add \$32 for each additional shipping address

### Payment and Credit Card Details

**Enclosed:** Personal Check \_\_\_\_\_  
Credit Card Payment Details \_\_\_\_\_  
Checks must be paid in U.S. dollars and drawn on a U.S. Bank.  
Credit Card:  VISA  Am. Exp.  MasterCard  
Card Number \_\_\_\_\_  
Expiration Date \_\_\_\_\_  
Signature: \_\_\_\_\_

Please send your order form and prepayment made payable to:

**Cadmus Reprints**  
**P.O. Box 751903**  
**Charlotte, NC 28275-1903**

*Note: Do not send express packages to this location, PO Box.  
FEIN #:541274108*

Signature \_\_\_\_\_ Date \_\_\_\_\_  
Signature is required. By signing this form, the author agrees to accept the responsibility for the payment of reprints and/or all charges described in this document.

### Invoice or Credit Card Information

#### Invoice Address Please Print Clearly

Please complete Invoice address as it appears on credit card statement

Name \_\_\_\_\_  
Institution \_\_\_\_\_  
Department \_\_\_\_\_  
Street \_\_\_\_\_  
City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_  
Country \_\_\_\_\_  
Phone \_\_\_\_\_ Fax \_\_\_\_\_  
E-mail Address \_\_\_\_\_

**Cadmus will process credit cards and Cadmus Journal  
Services will appear on the credit card statement.**

*If you don't mail your order form, you may fax it to 410-820-9765 with  
your credit card information.*



# Radiology 2008

## Black and White Reprint Prices

Domestic (USA only)						
# of Pages	50	100	200	300	400	500
1-4	\$221	\$233	\$268	\$285	\$303	\$323
5-8	\$355	\$382	\$432	\$466	\$510	\$544
9-12	\$466	\$513	\$595	\$652	\$714	\$775
13-16	\$576	\$640	\$749	\$830	\$912	\$995
17-20	\$694	\$775	\$906	\$1,017	\$1,117	\$1,220
21-24	\$809	\$906	\$1,071	\$1,200	\$1,321	\$1,471
25-28	\$928	\$1,041	\$1,242	\$1,390	\$1,544	\$1,688
29-32	\$1,042	\$1,178	\$1,403	\$1,568	\$1,751	\$1,924
Covers	\$97	\$118	\$215	\$323	\$442	\$555

## Color Reprint Prices

Domestic (USA only)						
# of Pages	50	100	200	300	400	500
1-4	\$223	\$239	\$352	\$473	\$597	\$719
5-8	\$349	\$401	\$601	\$849	\$1,099	\$1,349
9-12	\$486	\$517	\$852	\$1,232	\$1,609	\$1,992
13-16	\$615	\$651	\$1,105	\$1,609	\$2,117	\$2,624
17-20	\$759	\$787	\$1,357	\$1,997	\$2,626	\$3,260
21-24	\$897	\$924	\$1,611	\$2,376	\$3,135	\$3,905
25-28	\$1,033	\$1,071	\$1,873	\$2,757	\$3,650	\$4,536
29-32	\$1,175	\$1,208	\$2,122	\$3,138	\$4,162	\$5,180
Covers	\$97	\$118	\$215	\$323	\$442	\$555

International (includes Canada and Mexico)						
# of Pages	50	100	200	300	400	500
1-4	\$272	\$283	\$340	\$397	\$446	\$506
5-8	\$428	\$455	\$576	\$675	\$784	\$884
9-12	\$580	\$626	\$805	\$964	\$1,115	\$1,278
13-16	\$724	\$786	\$1,023	\$1,232	\$1,445	\$1,652
17-20	\$878	\$958	\$1,246	\$1,520	\$1,774	\$2,030
21-24	\$1,022	\$1,119	\$1,474	\$1,795	\$2,108	\$2,426
25-28	\$1,176	\$1,291	\$1,700	\$2,070	\$2,450	\$2,813
29-32	\$1,316	\$1,452	\$1,936	\$2,355	\$2,784	\$3,209
Covers	\$156	\$176	\$335	\$525	\$716	\$905

International (includes Canada and Mexico))						
# of Pages	50	100	200	300	400	500
1-4	\$278	\$290	\$424	\$586	\$741	\$904
5-8	\$429	\$472	\$746	\$1,058	\$1,374	\$1,690
9-12	\$604	\$629	\$1,061	\$1,545	\$2,011	\$2,494
13-16	\$766	\$797	\$1,378	\$2,013	\$2,647	\$3,280
17-20	\$945	\$972	\$1,698	\$2,499	\$3,282	\$4,069
21-24	\$1,110	\$1,139	\$2,015	\$2,970	\$3,921	\$4,873
25-28	\$1,290	\$1,321	\$2,333	\$3,437	\$4,556	\$5,661
29-32	\$1,455	\$1,482	\$2,652	\$3,924	\$5,193	\$6,462
Covers	\$156	\$176	\$335	\$525	\$716	\$905

Minimum order is 50 copies. For orders larger than 500 copies, please consult Cadmus Reprints at 800-407-9190.

### Reprint Cover

Cover prices are listed above. The cover will include the publication title, article title, and author name in black.

### Shipping

Shipping costs are included in the reprint prices. Domestic orders are shipped via UPS Ground service. Foreign orders are shipped via a proof of delivery air service.

### Multiple Shipments

Orders can be shipped to more than one location. Please be aware that it will cost \$32 for each additional location.

### Delivery

Your order will be shipped within 2 weeks of the journal print date. Allow extra time for delivery.

### Tax Due

Residents of Virginia, Maryland, Pennsylvania, and the District of Columbia are required to add the appropriate sales tax to each reprint order. For orders shipped to Canada, please add 7% Canadian GST unless exemption is claimed.

### Ordering

Reprint order forms and purchase order or prepayment is required to process your order. Please reference journal name and reprint number or manuscript number on any correspondence. You may use the reverse side of this form as a proforma invoice. Please return your order form and prepayment to:

**Cadmus Reprints**  
P.O. Box 751903  
Charlotte, NC 28275-1903

*Note: Do not send express packages to this location, PO Box. FEIN #: 541274108*

Please direct all inquiries to:

**Rose A. Baynard**  
800-407-9190 (toll free number)  
410-819-3966 (direct number)  
410-820-9765 (FAX number)  
[baynardr@cadmus.com](mailto:baynardr@cadmus.com) (e-mail)

**Reprint Order Forms and purchase order or prepayments must be received 72 hours after receipt of form.**