Physician Gender Effects in Medical Communication
A Meta-analytic Review

Debra L. Roter, DrPH
Judith A. Hall, PhD
Yutaka Aoki, MS, MHS, ME

STUDIES HAVE LINKED PHYSICIANS’ communication skills to a variety of positive outcomes, including patient and physician satisfaction, higher levels of adherence to therapeutic recommendations, improved physiological indicators of disease control, and enhanced physical and mental health status.1-3 Within this context, gender has stimulated a good deal of interest as a possible source of variation in the interpersonal aspects of medical practice, with speculation that female physicians facilitate more open and equal exchange and a different therapeutic milieu from that of male physicians.4-7

Outside of the medical context, differences in the interpersonal style of women compared with men are well documented.8-11 Women disclose more information about themselves in conversation,10 they have a warmer and more engaged style of nonverbal communication,8 and they encourage and facilitate others to talk to them more freely and in a warmer and more intimate way.8 There is also evidence that women take greater pains to downplay their own status in an attempt to equalize status with a partner, in contrast with men’s tendency to assert status differences.11 Despite gender differences in routine conversation, it is not known whether “female-linked” conversational styles are

Context  Physician gender has been viewed as a possible source of variation in the interpersonal aspects of medical practice, with speculation that female physicians facilitate more open and equal exchange and a different therapeutic milieu from that of male physicians. However, studies in this area are generally based on small samples, with conflicting results.

Objective  To systematically review and quantify the effect of physician gender on communication during medical visits.

Data Sources  Online database searches of English-language abstracts for the years 1967 to 2001 (MEDLINE, AIDSLINE, PsycINFO, and Bioethics); a hand search was conducted of reprint files and the reference sections of review articles and other publications.

Study Selection  Studies using a communication data source, such as audiotape, videotape, or direct observation, and large national or regional studies in which physician report was used to establish length of visit, were identified through bibliographic and computerized searches. Twenty-three observational studies and 3 large physician-report studies reported in 29 publications met inclusion criteria and were rated.

Data Extraction  The Cohen d was computed based on 2 reviewers’ (J.A.H. and Y.A.) independent extraction of quantitative information from the publications. Study heterogeneity was tested using Q statistics and pooled effect sizes were computed using the appropriate effects model. The characteristics of the study populations were also extracted.

Data Synthesis  Female physicians engage in significantly more active partnership behaviors, positive talk, psychosocial counseling, psychosocial question asking, and emotionally focused talk. There were no gender differences evident in the amount, quality, or manner of biomedical information giving or social conversation. Medical visits with female physicians are, on average, 2 minutes (10%) longer than those with male physicians. Obstetrics and gynecology may present a different pattern than that of primary care, with male physicians demonstrating higher levels of emotionally focused talk than their female colleagues.

Conclusions  Female primary care physicians engage in more communication that can be considered patient centered and have longer visits than their male colleagues. Limited studies exist outside of primary care, and gender-related practice patterns in some subspecialties may differ from those evident in primary care.
©2002 American Medical Association. All rights reserved.

PHYSICIAN GENDER EFFECTS IN MEDICAL COMMUNICATION

Box. Conceptual Groupings of Physician Communication Categories

Information giving
- Information content
- Biomedical
- Psychosocial
- Informational manner
- Directive
- Nondirective
- Information quality

Questions
- Question content
- General
- Biomedical
- Psychosocial
- Compliance
- Question format
- Closed
- Open

Partnership building
- Enlistment
- Lowered dominance

Length of visit
- Physician-reported time
- Observed time

Socioemotional behavior
- Social conversation
- Positive talk
- Negative talk
- Focus on emotions
- Positive nonverbal communication

Analytic Approach

A systematic review of the studies produced more than 150 different communication variables related to physician gender. Starting with a conceptual framework developed in an earlier meta-analysis evaluating correlates of physician communication in medical visits, the majority of variables were easily classified into several subsuming categories. The remaining variables were assigned to groupings based on an agreement between the first 2 authors, as reflected in the Box. (A listing of the individual variables, the study citations from which the variables were drawn, their assignment to each category, and a notation indicating the variables for which ≥1 studies showed a significant physician gender effect is available on request from the authors.)

For measures reflecting differences in length of visit, effect size was expressed as the mean difference between male and female physicians measured in minutes. For the other communication variables, however, the scale of measurement often varied across studies so that direct comparison of mean values across studies, even when applied to conceptually equivalent variables, was not useful. Consequently, gender differences were expressed as a Cohen’s d, defined as the difference between the male and female means divided by their pooled within-group SD. For this measure of effect size, a positive Cohen’s d value indicates that female physicians scored higher on the variable in question than male physicians did, and a negative Cohen’s d value indicates the reverse. In the studies summarized herein, the Cohen’s d was never reported directly but, rather, was calculated from the published information using standard formulas (eg, means and SDs, frequencies or percentages, correlation coefficients, or the χ² test, t test, or F test). Two investigators (J.A.H. and Y.A.) independently abstracted the quantitative information from the publications and calculated the Cohen’s d. If a study reported a result as nonsignificant and gave no other useful data for calculation of the Cohen’s d, a conservative approach was taken and a 0 was used in the calculation. A final consideration in the presentation of results is the magnitude of the effect size. Cohen’s rule of thumb for effect size considers 0.2 as small, 0.5 as medium, and 0.8 or greater as large.

When there was substantial heterogeneity (P > .10 for heterogeneity based on Q statistics) in the effect sizes within a given category, a random-effects model was used in the calculation of the Cohen’s d (pooled effect size). When the test for heterogeneity was nonsignificant, a fixed-effects model was used. In addition to the Cohen’s d, a standard normal deviate (z score, the statistic associated with a P value; eg, the z score associated with P = .05 by a 2-tailed test is 1.96) was derived for each result, and these were added and divided by the square root of the total number of studies to obtain a z score and an associated combined P value (Stouffer method). If an

©2002 American Medical Association. All rights reserved.

Method

Search Procedure and Criteria for Study Inclusion

To be included in the review, a study had to (1) involve physicians, physicians in training (interns or residents), or medical students; (2) involve actual or standardized patients; (3) measure communication using neutral observers (including standardized patients as observers), audiotape, or videotape, with the exception of inclusion of physician-reported length of the medical visit; (4) test for an association between physician gender and at least 1 interpretable physician communication variable; (5) deal with nonpsychiatric medical visits; and (6) be published in an English-language book or journal.

Studies were identified through online database searches of MEDLINE (1967-2001), AIDSLINE, PsycINFO, and Bioethics using the key words doctor-patient interaction, patient-interaction, physician-patient interaction, and doctor-patient relationship. These key words were combined with other key words: female, gender effects, female physicians, female doctors, and effect of sex of doctor. In addition, a hand search was conducted of our own reprints files and the reference sections of review articles and other publications.

Methods

For measures reflecting differences in length of visit, effect size was expressed as the mean difference between male and female physicians measured in minutes. For the other communication variables, however, the scale of measurement often varied across studies so that direct comparison of mean values across studies, even when applied to conceptually equivalent variables, was not useful. Consequently, gender differences were expressed as a Cohen’s d, defined as the difference between the male and female means divided by their pooled within-group SD. For this measure of effect size, a positive Cohen’s d value indicates that female physicians scored higher on the variable in question than male physicians did, and a negative Cohen’s d value indicates the reverse. In the studies summarized herein, the Cohen’s d was never reported directly but, rather, was calculated from the published information using standard formulas (eg, means and SDs, frequencies or percentages, correlation coefficients, or the χ² test, t test, or F test). Two investigators (J.A.H. and Y.A.) independently abstracted the quantitative information from the publications and calculated the Cohen’s d. If a study reported a result as nonsignificant and gave no other useful data for calculation of the Cohen’s d, a conservative approach was taken and a 0 was used in the calculation. A final consideration in the presentation of results is the magnitude of the effect size. Cohen’s rule of thumb for effect size considers 0.2 as small, 0.5 as medium, and 0.8 or greater as large.

When there was substantial heterogeneity (P > .10 for heterogeneity based on Q statistics) in the effect sizes within a given category, a random-effects model was used in the calculation of the Cohen’s d (pooled effect size). When the test for heterogeneity was nonsignificant, a fixed-effects model was used. In addition to the Cohen’s d, a standard normal deviate (z score, the statistic associated with a P value; eg, the z score associated with P = .05 by a 2-tailed test is 1.96) was derived for each result, and these were added and divided by the square root of the total number of studies to obtain a z score and an associated combined P value (Stouffer method). If an
Table 1. Number of Physicians and Studies by Physician Specialty and Training Level

<table>
<thead>
<tr>
<th>Specialty</th>
<th>No. of Male Physicians</th>
<th>No. of Female Physicians</th>
<th>No. of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal medicine</td>
<td>164</td>
<td>72</td>
<td>4</td>
</tr>
<tr>
<td>Family practice</td>
<td>97</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>General practice</td>
<td>81</td>
<td>56</td>
<td>5</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>39</td>
<td>34</td>
<td>2</td>
</tr>
<tr>
<td>Obstetrics/gynecology</td>
<td>24</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>Internal medicine/family practice</td>
<td>79</td>
<td>68</td>
<td>3</td>
</tr>
<tr>
<td>No stated specialty</td>
<td>81</td>
<td>52</td>
<td>3</td>
</tr>
<tr>
<td>Physician training status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical student†</td>
<td>81</td>
<td>52</td>
<td>3</td>
</tr>
<tr>
<td>Resident</td>
<td>206</td>
<td>141</td>
<td>9</td>
</tr>
<tr>
<td>Practitioner</td>
<td>91</td>
<td>58</td>
<td>7</td>
</tr>
<tr>
<td>All levels</td>
<td>87</td>
<td>94</td>
<td>4</td>
</tr>
</tbody>
</table>

*Data are from observational studies only.
†Data not available by gender breakdown for 1 of these 3 studies.

Table 2. Mean Number (Range) of Participants and Visits per Study

<table>
<thead>
<tr>
<th>Study Type</th>
<th>Participants per Study</th>
<th>Visits per Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observational Studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>25 (5-101)</td>
<td>97 (16-196)</td>
</tr>
<tr>
<td>Female</td>
<td>15 (3-33)</td>
<td>66 (4-164)</td>
</tr>
<tr>
<td>Reported Studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>665 (27-1836)</td>
<td>22,367 (442-48,705)</td>
</tr>
<tr>
<td>Female</td>
<td>91 (23-154)</td>
<td>10,662 (794-51,044)</td>
</tr>
</tbody>
</table>

The author reported a result as nonsignificant and gave no other useful data for calculation of a z score, a z of 0 was used in the calculation of the combined probability. Calculating a combined value that includes all studies captures information that is often embedded in null results and generally lost, and provides a commonly understood probability measure to compare results across variables of interest. In this analysis, some studies provided more than 1 result for a given communication category. For example, a hypothetical study might measure “shared decision making,” “lets patient state concerns,” and “encourages patient paraphrasing,” all of which would have been placed in the category of partnership building. Because of conceptual redundancy, only the strongest single association for a given category in a given study was entered in summaries of effect size and z score.

STATA statistical software and user-written modules were used for all analyses. Throughout this article, P < .05 is considered statistically significant.

Descriptive Summary of the Literature

Twenty-six studies, reported in 29 publications, were included in the meta-analytic calculations; 23 of these were direct observation studies and 3 were based on physician report of visit length. (References 27 and 28 are publications reporting data from the same study. This is also the case for references 43 and 44 as well as 45 and 46.)

As reflected in Table 1, most studies were conducted in primary care settings, and physicians at all levels of training were represented in the studies. The average number of physicians in the observational studies was 40, with male physicians substantially outnumbering female physicians (n = 25 and 15, respectively; Table 2). The average number of visits per study was 157; this reflected an average of 97 visits to male physicians and 65 visits to female physicians in each study. There was wide variation in the number of patients observed by each physician; the average was 4, with a range of 1 to 32.

Three large databases were used to supplement the observational data on length of visit by physician report. The US analysis was based on the National Ambulatory Medical Care Survey and was used to derive separate estimates of physician gender effects for length of visit in 5 specialties: general and family practice, internal medicine, pediatrics, obstetrics and gynecology, and dermatology. A similar Dutch analysis derived estimates of length of visit from the Dutch National Study on Morbidity and Interventions in General Practice for general practitioners, with a subanalysis of general practitioners treating patients for “women’s health problems.” The third report represents a regional US study.

The majority of studies were based on complete patient visits; however, 3 studies limited evaluation to a specific portion of the visit and 6 studies were of a standardized patient interview. Ten of the studies were based on all-female patient populations, which included mothers with their children seeking pediatric care (children were of both sexes), standardized patients (all of whom were female), and women seeking obstetrical care. There were no studies of exclusively male patient populations.

Eighteen different systems were used to measure communication variables. Only 2 of these systems were used in more than 1 study; the Beckman and Frankel method for assessment of physician probing of patient concerns was used in 2 studies and the Roter Interaction Analysis System was used in 3 studies. Eighteen of the 23 observational studies reported intercoder reliability; all were in the mid to high range (0.59-1.0) based on Pearson correlation coefficients or κ statistics.

RESULTS

Details of the results reviewed below relating effects of physician gender to categories of patient-physician communication are displayed in Table 3. The Figure summarizes these results graphically.

Information Giving

Informational Content. Nine studies addressed the relationship of biomedical information giving and physician gender. Two of these studies reported
significant results; one reported significantly higher levels for male physicians, but the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.

Informational Manner. Information can be given in a more or less directive manner; for instance, giving specific instructions is an example of the former, while proposing alternatives is an example of the latter. Nine studies produced significant results; one reported significantly higher levels for male physicians, the other reported the opposite. Neither the combined z score nor the Cohen d was statistically significant.

A more consistent picture of gender effect emerged for psychosocial discussion. Five of the 10 studies addressing psychosocial discussion reported significantly higher levels for female rather than male physicians; only 1 study (a study of gynecologists) reported higher (but nonsignificant) levels of psychosocial discussion for male physicians. Both the combined z score and the pooled Cohen d were significant, as shown in Table 3.
ies assessed variables that could be considered directive; 2 of these found higher levels of directiveness for male physicians and 1 reported higher levels for female pediatricians. Five studies assessed variables consistent with nondirective information giving. Of these, only 1 study reported a significant result indicating higher levels for female physicians. Neither the overall combined z score summarizing these studies nor the pooled effect size was significant.

Informational Quality. Measures of information-giving quality include such variables as clear communication and avoidance of jargon. Six studies investigated this issue and none reported statistically significant results.

Question Content and Form
Questions are broadly concerned with data gathering and may reflect general question asking (simply defined as requesting information) or content-specific inquiries (questions with biomedical, psychosocial, compliance-related content). Questions were sometimes also specified by form, either as open-ended or closed-ended. (Questions with the intent of furthering patient introspection or clarifying patient expectations were classified as partnership building and are discussed later.)

There was little evidence of a gender effect for general question asking, for biomedical questions, or for compliance-related questions. Five studies identified questions generally; 2 of the 3 studies that reported significant results reported significantly higher levels of question asking among female physicians, while another found the opposite, yielding a nonsignificant pooled effect. For biomedical questions, 2 studies reported significant results; one reported significantly higher levels of question asking among male physicians, while the other found the opposite, with a nonsignificant pooled effect. Two studies specified questions related to patient compliance and neither reported a significant gender effect (results not shown).

As was evident in the content analysis of information giving, women showed higher levels of psychosocial question asking than men; 3 of 6 studies reported significant results indicating higher levels for female physicians. No studies reported higher levels of psychosocial questioning by male physicians. The combined z score was statistically significant and the Cohen d showed a trend toward significance (P = .06).

One of 4 studies that assessed closed questions reported significantly higher levels among women and none reported higher levels for men. The pooled findings reflect a statistically significant Cohen d and a trend toward a significant combined z score (P = .09). In contrast, we found little evidence of gender effect on open-ended questions; 1 study of 6 found significantly higher levels for female physicians, but neither the combined z score nor the Cohen d was statistically significant.

Partnership Building
By our definition, partnership building occurs when a physician actively facilitates patient participation in the medical visit or attempts to equalize status by assuming a less dominating stance within the relationship. The 2 classes of partnering behavior can be distinguished as reflecting (1) an active facilitation of partnership, or “enlistment,” or (2) a more passive or “lowered-dominance” approach to partnership (see the Box for examples). Twelve studies included the active, enlistment-type variables in their assessments. Six of these studies reported significantly higher levels of enlistment on the part of female physicians, while studies showed the reverse. Heterogeneity analysis indicated substantial variation among the studies. Both the combined z score and the pooled Cohen d were statistically significant, as shown in Table 3. (We became aware late in this article’s production of an overlooked study that examined communication elements that would have been included in the active partnership category. The study assessed facilitative phrases with question tags such as “isn’t it?” and “don’t you?” The study reported significantly higher use of the question tag “don’t you?” by female compared with male physicians [P = .03; Cohen d = 0.32]. The question tag “isn’t it?” did not show a significant gender effect.)

Five studies assessed variables reflecting the passive, lowered-dominance approach and 1 of these reported a statistically significant result indicating lowered dominance among female physicians. Neither the Cohen d nor the combined z score summarizing these studies was significant.

Social Conversation
Social conversation is defined as nonmedical exchanges, largely social pleasantries and greetings. Seven studies included measures of social conversation. One study reported a statistically significant result reflecting higher levels of social conversation by female pediatricians. The measures summarizing these studies were not significant, as shown in Table 3.

Positive Talk
Positive talk assessment captures the generally positive atmosphere created in the visit through verbal behaviors such as agreements, encouragement, and reassurance. Social conversation was not included in this category, with the exception of 2 studies in which it was embedded in a composite variable composed of positive elements. Fourteen studies included some measure of positive talk. Six of these studies reported significantly higher levels of positive talk for female physicians, while 2 studies showed the reverse. Seven studies reported significantly higher levels of positive talk for male physicians. The Cohen d was significant and the combined z score summarizing these studies was large, as shown in Table 3.

Eight of the studies measuring positive talk were based on exclusively female patient populations, while 6 studies were based on both male and female patients. Inspection of the pattern of results (data not shown) shows evidence of a physician gender effect for...
all-female patient populations (pooled effect size, 0.25; 95% confidence interval [CI], −0.01 to 0.52; \( P = .06 \)) as well as mixed-gender populations (pooled effect size, 0.59; 95% CI, 0.41−0.77; \( P < .001 \)). Comparison of the pooled effect size estimates from these 2 populations suggests a significantly greater effect in mixed-gender populations than in exclusively female populations (\( P = .01 \)).

### Negative Talk

The negative talk category included explicit verbal expressions of criticism or disapproval. Nine studies assessed a significant element of negative talk. Two of these studies reported a significant result, one of which indicated that female physicians were more likely to avoid being critical of a patient than male physicians and the other of which reflected higher levels of disagreement by female compared with male obstetricians. Neither the Cohen \( d \) nor the overall combined \( z \) score summarizing these studies was significant, as shown in Table 3.

### Emotionally Focused Talk

Emotionally focused talk included explicit inquiry about feelings and emotions, exploration of emotional concerns, and statements of empathy and concern. This category is distinguished from psychosocial exchange by its direct link to feelings and emotions. Thirteen studies assessed emotional talk in some manner; 4 of these found significantly higher levels for female compared with male physicians. Both gynecology studies in the review found higher levels of emotional talk for male physicians; one of these reported a significant result and the other was marginally significant (\( P = .06 \)). The combined \( z \) score was statistically significant but the pooled Cohen \( d \) was not, as shown in Table 3.

The studies showed a high degree of heterogeneity, which was almost entirely explained by the 2 obstetrics and gynecology studies. Estimates based on the pooling of the 11 primary care studies showed a significant and consistent gender effect favoring female physicians (effect size, 0.16; 95% CI, 0.04−0.29; \( P = .008 \)), with a low degree of heterogeneity (\( P \) for heterogeneity = .19). Analysis of the 2 obstetrics and gynecology studies showed a significant gender effect favoring male physicians (effect size, −0.27; 95% CI, −0.49 to −0.07; \( P = .01 \)), with a low degree of heterogeneity (\( P = .41 \) for heterogeneity).

### Nonverbal Communication

Nonverbal communication includes positive nonverbal behaviors (smiles, nods, friendly voice tone, relaxed hands), and a variety of behaviors that can have ambiguous, neutral, or negative meaning depending on the context of use (eg, touches patient, folds hands, gestures while speaking, points at the patient, speech is disturbed, voice tone reflects anxiety or boredom). To avoid the difficulties associated with ambiguous interpretation, we limited our analysis to positive nonverbal behaviors.

Six studies assessed positive nonverbal behavior in some manner, and 2 of these studies reported significant results showing that female physicians demonstrate higher levels of smiling and head nods and awareness of non-verbal communication. No studies reported higher levels of positive nonverbal behavior for male physicians. The combined \( z \) score was significant (\( P = .02 \)) and the Cohen \( d \) trended toward significance (\( P = .07 \)), as shown in Table 3.

### Length of Visit

Five of the 10 studies that directly measured length of visit reported women to conduct significantly longer visits than men and only 1 found statistically longer visits for male physicians (a study of obstetricians). Length of visit averaged 21 minutes (range, 7.4−36.7 minutes) for male physicians and 23 minutes (range, 10.5−37.0 minutes) for female physicians. The 3 studies of physician-reported length of visit similarly found that female physicians’ visits were significantly longer than male physicians’ visits, with an average of 14.8 minutes for male physicians and 17.0 minutes for female physicians. (Visits in the Netherlands are substantially shorter than visits in the United States, regardless of physician gender.)

There was no statistically significant difference between the observational and physician-reported estimates of length of visit attributed to physician gender; based on the 10 observational studies, visits with female physicians averaged 2.0 minutes longer than visits with male physicians (95% CI, 0.5−3.5) and were 2.1 minutes longer (95% CI, −0.15 to 4.3) based on the 7 physician-reported studies. The overall difference in length of visit attributed to physician gender for all 17 studies was 2.05 minutes (95% CI, 0.65−3.4).

### Comment

Despite widespread interest in the effects of physician gender on the care process, the literature describing these effects is small. We identified only 23 observational studies relating the communication process to physician gender. Nevertheless, the pattern of results was almost entirely consistent with what one might expect from the non-medical literature regarding gender differences in communication. Female physicians engage in communication that more broadly relates to the larger life context of patients’ conditions by addressing psychosocial issues through related questioning and counseling, greater use of emotional talk, more positive talk, and more active enlistment of patient input. When taken together, these elements comprise a pattern that can be broadly considered “patient-centered” interviewing. In contrast with the higher levels of psychosocial and socioemotional exchange, there is little evidence that physician gender is related to the more task-specific communication elements of care. Physician gender was not related to the provision of biomedical information (including discussion of the diagnosis, prognosis, and medical treat-
Behavioral differences in the communication styles of male and female physicians would be especially important if they produce corresponding gender differences in patients’ behavior directed back at them. Indeed, a separate meta-analysis investigating the effects of physician gender on patient communication suggested that patient behavior largely reciprocates the gender-linked physician behavior reported in the current analysis,93 like female physicians, their patients talk more overall, make more positive statements, discuss more psychosocial information, and express more partnership building than patients of male physicians. Patient behaviors that showed little or no difference in relation to physician gender were patient questions, social conversation, and negative statements. Some behaviors were indirectly reciprocated. Even though male and female physicians did not differ in how much biomedical information they provided to their patients, patients of female physicians provided more biomedical information to them than to male physicians. Since female physicians ask more psychosocial questions than their male counterparts, it may be that this type of question stimulates more patient disclosure of both a psychosocial and biomedical nature. More patient disclosure may also be fostered by female physicians’ more active efforts to build partnership through inviting patients’ opinions and through the use of facilitators and back channels, such as saying “uh-huh” and “go on,” and other indicators of verbal attentiveness. Interestingly, though female physicians made more emotionally focused statements than male physicians, patients did not direct more emotional discussion to their female compared with male physicians. Patient and physician gender concordance appears to strengthen many of the gender effects observed. The 2 studies of which we are aware that have investigated the impact of gender concordance on communication27,28 both found that female gender concordant visits were characterized by longer length and more equal patient and physician contributions to the medical dialogue than all other gender combinations. Hall et al27 also reported more positive statements, nodding, and back channels in female-concordant visits compared with others.

Female physicians also spent more time with their patients than male physicians did, an average difference of about 2 minutes, or 10%, per visit. With the increasing time and productivity pressures that plague all physicians, a 2-minute-per-visit increase represents a substantial burden, easily putting a female physician an hour behind her male colleagues at the end of a busy day. Mechanic et al94 have reported that the average medical visit has increased by 1 to 2 minutes in the last 10 years. Despite this increase, the widespread perception of a shrinking visit may be fueled by the time-pressured atmosphere within which physicians may be providing more preventive and counseling services than in the past.54-56 In this light, female physicians may be at even greater risk of falling behind their male colleagues in daily scheduling. The analysis by Henderson and Weissman57 of the Commonwealth survey of patient-reported screening and counseling services concluded that female physicians provided more preventive counseling to both their male and female patients, and more gender-specific screening to their female patients than did male physicians.

Pressure to do more in limited time may act to further amplify the communication differences between physicians of different genders. While male physicians may respond to time pressures by dispensing with socioemotional and psychosocial tasks, as suggested by Mechanic,90 female physicians may find this more difficult to do.90 We suggest this because female physicians currently record proportionately more diagnoses of a psychosocial nature than their male colleagues do,26,44 and the demand for diagnosis and treatment of mental health problems in primary care is expected to grow.50

The results from the 2 obstetrics and gynecology studies deviate from those of the primary care studies. As several studies have documented especially strong patient preferences for female physicians in gynecologic and obstetric care,63,62 male physicians may feel pressured to meet the increasing competitive challenge of growing numbers of female physicians by enhancing their interpersonal skills.61 If this is the case, it would suggest that physicians are capable of modifying their communication style given sufficient motivation and incentive. The training literature is also optimistic in this regard; there is ample evidence that instruction in communication skills is associated with improvement in skills, with some studies showing these improvements to be long-lasting64-66.

The literature on which the current analysis depends is small and limited to published studies raising the possibility that relevant, unpublished material was missed. However, the fairly high number of null results in the review indicates that many studies lacking statistically significant findings were published. This is consistent with the fact that gender effects were often of secondary and even tertiary importance in terms of the main theme of an article, making it unlikely that unpublished studies would have produced a different pattern of results.27 Furthermore, we have no reason to believe that results favoring one gender over the other would be preferentially published. It is also noteworthy that more than half of the communication measures showed significant heterogeneity across studies. While we do not know what the source of unmeasured variability may be, it is likely to be some aspect of the institutional cultures of the study sites, as well as case mix and delivery system.49,54

What can we conclude about the consequences of these gender-related communication effects in terms of the variety of patient outcomes so valued in health services research? The reviewed
studies did not systematically address patient outcomes, and no direct conclusion can be drawn. It seems likely that the effects found are an indication of a relatively more health-promoting therapeutic milieu produced by female physicians. Such a conclusion, however, can only be speculative, since no study has directly investigated whether patients of female physicians fare better on clinical measures. Furthermore, while physician communication behaviors similar to those reviewed here have been positively related to patient satisfaction, compliance, and patient recall and comprehension of information, as indicated by a review of studies, it is not always the case. On average, female physicians fare better on clinical measures. Some show more satisfaction with female physicians, some show more satisfaction with female physicians, and some show no difference. It is premature to offer an explanation for this variation, and we can only speculate that patients’ satisfaction depends on what the physician actually does, as well as stereotypes and expectations held by patients and differences in patient characteristics, such as health status or sociodemographics. Future studies of physician gender and communication will need to focus greater attention on the assessment of patient health outcomes and other indices of care quality.

In summary, the review produced a pattern of effects associated with physician gender that goes beyond a listing of individual elements of medical exchange. Taken together, the differences reflect a patient-centered communication style that inspires patient reciprocation and is likely to reflect a more intimate therapeutic milieu of heightened engagement, comfort, and partnership. Although the magnitude of the effects attributable to gender for any given communication element were usually small, the effects are comparable with those of many well-established medical, psychological, behavioral, and educational interventions. What might these results mean for male physicians? We do not suggest that all or even most female physicians are patient-centered and male physicians are not; there is far more common ground than difference in the communication behaviors of male and female physicians. Moreover, physicians, both male and female, who are skillful communicators may achieve time efficiencies that allow the delivery of quality, patient-centered care in even restricted time frames. Physicians have the capacity to improve their communication skills in meaningful ways through self-awareness, self-monitoring, and training. The potentially powerful impact of patient reciprocation of both communication style and affect in the medical visit is especially important to recognize, as recognition could help create positive exchanges and defuse negatively spiraling interaction patterns. Both male and female physicians, as well as their patients, are entitled to no less.

Author Contributions: Study design and concept: Roter, Hall. Acquisition of data: Roter. Analysis and interpretation of data: Roter, Hall, Aoki. Drafting of the manuscript: Roter, Hall. Critical revision of the manuscript for important intellectual content: Roter, Hall, Aoki. Statistical expertise: Roter, Hall, Aoki. Administrative, technical, or material support: Roter. Study supervision: Roter.

REFERENCES
PHYSICIAN GENDER EFFECTS IN MEDICAL COMMUNICATION

45. van den Brink-Muinen A, Bensing JM, Kerssens JJ. Gender and communication style in general practice: differences between women’s health care and regular health care. Med Care. 1998;36:100-106.
46. van den Brink-Muinen A. Gender. Health and Health Care in General Practice. Utrecht, the Netherlands: Nivel; 1996.