An Observational Study of Antibiotic Prescribing Behavior and the Hawthorne Effect

Rita Mangione-Smith, Marc N. Elliott, Laurie McDonald, and Elizabeth A. McGlynn

Objectives. To assess whether prospective, observational study procedures, including questionnaires and audio recording, are associated with different patterns of physician diagnostic decision making and antibiotic prescribing.

Data Sources/Setting. (1) Survey data from a prospective observational study of treatment patterns for children with acute upper respiratory illnesses (10/96–3/97) and (2) retrospective medical record abstraction data of nonobserved encounters for the same problems occurring during (10/96–3/97) and one year after (10/97–3/98) the observational study period. Ten pediatricians in two community practices were studied.

Study Design. Patterns of diagnoses recorded in the medical record and antibiotics ordered for visits occurring outside of the observational study (same time period and one year later) were compared with the pattern of diagnoses and antibiotics ordered during the observational study.

Data Collection/Extraction Methods. For the observational study (10/96–2/97), diagnosis and treatment choices were obtained from questionnaires completed by physicians immediately following the visit. For the nonstudy encounters (10/96–3/97 and 10/97–3/98), data were abstracted from medical records one year after the observational study was completed.

Principal Findings. The proportion of viral cases in which an antibiotic was prescribed was 29 percentage points lower for the observational study compared to the retrospective analysis ($p < .05$). In one of two study sites, the proportion of cases assigned a bacterial diagnosis was 29 percentage points lower in the observational study period compared to the retrospective study ($p < .05$).

Conclusions. Observational study procedures including questionnaires and audio recording can affect antibiotic prescribing behavior. Future observational studies aimed at examining the frequency of inappropriate antibiotic prescribing should measure and adjust for the Hawthorne effect; without such adjustments, the results will likely underestimate the true degree of the problem. Future interventions aimed at decreasing inappropriate antibiotic prescribing should consider "harnessing" the Hawthorne effect through performance feedback to participating physicians.

Key Words. Antibiotic prescribing, Hawthorne effect, pediatrics

THE HAWTHORNE EFFECT

The Hawthorne effect refers to a phenomenon where a study subject's behavior and/or study outcomes are altered as a result of the subject's awareness of being under observation. This phenomenon was originally identified at the Hawthorne Works Plant of the Western Electric Company in Chicago (Roethlisberger and Dickson 1939). Several studies were conducted at this plant between the years 1924 and 1932 in order to identify working conditions that would increase the productivity of the personnel employed by the plant. The investigators found that worker productivity increased regardless of working conditions when the workers knew they were under observation. For example, both more light and less light in the workroom resulted in improved performance when workers were aware that their productivity was being measured. For research studies examining physician performance, the Hawthorne effect may be a significant explanatory factor in observed improvements. In fact, several interventions aimed at improving physician performance have deliberately included physician feedback and goal-setting as components of successful quality improvement interventions (Davis et al. 1995; Greco and Eisenberg 1993; Lohr, Brook, and Kaufman 1980). For studies examining physician performance but not including such factors, measuring the Hawthorne effect is necessary to isolate the true effects of the intervention itself.

Previous investigations demonstrating significant Hawthorne effects on study outcomes have usually been designed to improve performance or outcomes and the study participants were aware of the outcomes being measured (De Amici et al. 2000; Arborelius and Timka 1990; Carabin et al. 1999; Grufferman 1999). It has been suggested that the Hawthorne effect may have a significant impact only when there is a perceived demand for improved performance or outcomes (Arborelius and Timka 1990). However, some studies have found that simply monitoring a particular outcome can change study participant behavior if they know the investigators are interested in that outcome (De Amici et al. 2000). For example, Carabin et al. (1999) found that

Address correspondence to Rita Mangione-Smith, M.D., M.P.H., Assistant Professor, UCLA, Department of Pediatrics, 12-358 MDCC, 10833 LeConte Ave., Los Angeles, CA 90095-1752. All with RAND in Santa Monica, CA, are Marc N. Elliott, Ph.D., Statistician, Laurie McDonald, M.S., Senior Programmer, and Elizabeth A. McGlynn, Ph.D., Director of the Center for Research on Quality in Health Care.

having day care workers record the number of children absent because of diarrheal illness over a nine-month period decreased the coliform bacterial counts on the hands of the classroom teachers and children (one of several outcomes that were monitored). De Amici et al. informed patients in the preoperative period that they were part of a research study and would thus be closely monitored postoperatively for complications and pain levels (De Amici et al. 2000). These patients were significantly less likely to report that they had pain and had significantly higher scores on a measure of psychological well-being postoperatively when compared to a control group of patients who were simply consented for surgery in the regular manner.

DETERMINANTS OF ANTIBIOTIC PRESCRIBING

The decision to prescribe antibiotics is complex and involves multiple factors. Among these are the patient's age (Schwartz et al. 1997), the duration and worsening of symptoms (Davy, Dick, and Munk 1998), physical examination findings (Butler et al. 1998; Davy, Dick, and Munk 1998; Dosh et al. 2000; Le Saux, Pham et al. 1999), perceived parental expectations for antibiotics (Mangione-Smith et al. 1999; Vinson and Lutz 1993; Watson et al. 1999), concerns related to maintaining a positive doctor–patient/parent relationship (Butler et al. 1998), the parent's need to return to work (lack of "sick day care" available for their child) (Barden et al. 1998), concern about adverse outcomes if treatment is withheld (Butler et al. 1998; Dosh et al. 2000), physician demographics, and physician specialty (Mainous, Hueston, and Love 1998).

We conducted an observational study from October 1996 to April 1997 (Mangione-Smith et al. 1999) to assess how parent pre-visit expectations and physician perceptions of those expectations affected physician antibiotic prescribing patterns. This study involved a pre- and postvisit survey of parents, a postvisit survey of physicians, and audiotaping of the physician–parent encounter. To be eligible for the study, parents had to present with a child who was being seen because of cold symptoms. With permission from the University of California, Los Angeles, Institutional Review Board, we withheld information regarding our main objectives from participating physicians until data collection was completed. At the study's conclusion, physicians were debriefed regarding our specific aims. When recruiting physicians, we told them we were interested in measuring parental pre-visit expectations for acute care visits for their child, how these expectations affected doctor–parent communication, and whether parents were satisfied with their visit. Although

we attempted to conceal our interest in parent expectations as they related to antibiotic overprescribing, the participating physicians may have ascertained our study purpose given the prominence with which this topic is discussed in the pediatric research literature as well as the lay press.

Adjusted results from our observational study controlling for many of these potential determinants of antibiotic prescribing revealed that the only significant predictor of prescribing antibiotics for a presumed viral illness was the physician's perception that the parent expected to receive antibiotics. For children with presumed viral illnesses, when physicians thought parents expected antibiotics, they prescribed them 62 percent of the time (versus 7 percent when they did not think antibiotics were desired [$p = .02$]). When physicians thought the parent expected an antibiotic, they were also significantly more likely to give a bacterial diagnosis (70 percent versus 31 percent, $p = .04$) (Mangione-Smith et al. 1999).

Physicians in our observational study prescribed antibiotics for presumed viral diagnoses in 17 percent of cases. This rate is remarkably lower than the rate reported in 1996 for a national probability sample of pediatricians where the investigators found that antibiotics were prescribed 38 percent of the time for the common cold (Nyquist et al. 1998a). This finding raised the question as to whether our study field procedures had resulted in a Hawthorne effect on antibiotic prescribing rates for viral conditions.

To our knowledge, no studies of the Hawthorne effect have examined whether subjects in noninterventional observational studies, who are not directly informed about the outcomes of interest, change their behavior during the observation period with regard to those outcomes. The current study was conducted after completion of the original observational study in order to assess the degree to which the Hawthorne effect might have influenced the main outcomes measured in this observational study, that is, antibiotic prescribing for viral illnesses and the proportion of cases assigned bacterial diagnoses. This information is needed to assess the validity of data obtained using observational study designs that examine physician behaviors. If an intervention to change antibiotic prescribing patterns is developed based on the results of such observational studies, the validity of these data must be established.

METHODS

For the observational study, eligible pediatricians were initially contacted by phone to assess whether they were interested in participating in a study that

focused on issues of doctor–parent communication during acute care visits for children in their practice. If the physicians in the practice were interested, the research team went to the office and gave a slide presentation outlining the main objectives and field procedures for the study.

For each physician who agreed to participate, we collected data during three separate two-week periods occurring between 10/96 and 3/97. Each of these two-week data collection periods was separated by a two- to four-week time gap. This was done in an effort to more evenly distribute each physician's study visits over the entire study period.

Physicians who agreed to participate completed a postvisit survey for each study encounter that included three checklists: one for physical examination findings, one for diagnoses, and one for treatments prescribed or recommended (see Appendix 2). For diagnoses and treatments, physicians were also given the option to write in a diagnosis or therapy if they did not want to select one of the choices on the lists provided. Physicians in our study wrote in alternate diagnoses in 11 percent of cases. None of these cases represented either a bacterial or viral upper respiratory tract infection but rather represented diagnoses that were not eligible for study inclusion, such as cerumen impaction and acute gastroenteritis.

After the observational study was completed, the study team returned to the participating offices and presented a second slide presentation that outlined the key findings of the study and revealed our "true" objectives in conducting the study. During these poststudy presentations, physicians were asked if they had ascertained that we were interested in antibiotic prescribing during data collection.

For the nonobservational study, we abstracted medical records (see Appendix 1) for nonobservational study patients of the physicians who participated in the observational study in order to assess the degree to which the physicians' prescribing and diagnostic patterns changed while they were under observation. All medical record abstractions were performed after the observational study was completed. For each physician who participated in the observational study, we attempted to abstract the same number of encounters as the physician completed for the observational study. For example, if a physician completed 20 encounters in the observational study, we abstracted 20 medical record encounters for that physician for the nonobservational study. For some of the physicians, 5 to 10 more encounters were abstracted for the nonobservational study than were completed in the observational study (Table 1). We abstracted approximately half of the encounters for each physician from the same time period as the observational

Table 1:   Diagnoses Assigned During the Observational Study and
Nonobservational Study

Recorded Diagnosis	Observational Study	Nonobservational Study: Period 1 (10/96–3/97)	Nonobservational Study: Period 2 (10/97–3/98)
Viral	149 (55%)α	91 (55%)	77 (56%)
Bacterial	123 (45%)	76 (45%)	60 (44%)
Total	272	167	137

αNumber (%) of encounters

study data collection (10/96–3/97) but only included visit dates that fell
outside the dates of data collection for the observational study in that
physician's office. For example, if we had collected observational study data
during the last two weeks of 11/96 for a given physician, all eligible medical
record encounter dates for that two-week period would be excluded for that
physician. For all but one of the study physicians the other half of the
abstracted medical record encounters came from visits that occurred one year
after the observational study had been completed (10/97–3/98). One
physician left practice in 5/97, thus all abstracted medical record encounters
for this physician came from the first period (10/96–3/97). We excluded
charts of patients who were involved in the observational study from the
nonobservational study.

For each month of the nonobservational study (10/96–3/97 or 10/97–
3/98), one of two processes was used to identify medical records that
contained eligible visits for data abstraction. A visit was considered eligible for
abstraction if it was conducted by a study physician, the child's age fell
between 2 and 10 years at the time of the encounter, the visit occurred during
the appropriate time frame, and the child had one of the following diagnoses:
acute otitis media, otitis media with effusion, otitis externa, asthma, bronchitis,
pneumonia, mycoplasma infection, croup, streptococcal pharyngitis, viral
pharyngitis, viral upper respiratory infection (URI), sinusitis, or viral
syndrome. In one practice, administrative data were used to identify medical
records that contained visits that were eligible for data abstraction. In the
second practice, medical records were sequentially pulled and screened for
study eligibility by the first author (RM-S). Medical records in this latter
practice were arranged in alphabetical order. Because of administrative
constraints placed on us by the participating office we were unable to
randomly select charts for potential abstraction. Although selecting charts in
alphabetical order may have introduced selection bias based on race/
ethnicity, in our observational study we found no relationship between race/

ethnicity and antibiotic prescribing patterns or the diagnoses assigned. Additionally, arriving at the final sample for this practice required working through the entire alphabet. All encounters in a selected medical record were sequentially reviewed until we identified one that met the eligibility criteria for inclusion. For each medical record abstracted, only the first eligible encounter was included. If a medical record had no encounters meeting eligibility criteria, it was not included and the next chart in the sequence was pulled and reviewed. This process was repeated until the requisite number of encounters had been abstracted for each participating physician in the second practice.

The abstraction form collected the following data for each eligible encounter: (1) The child's age, (2) date of the encounter, (3) the diagnosis, (4) whether or not a chest x-ray, sinus x-ray, rapid strep test, throat culture, or complete blood count (CBC) was performed, and if so, the results of the test, and (5) whether or not an antibiotic was prescribed, and if so, the name of the antibiotic prescribed (see Appendix 1). The encounter was the unit of analysis for all study outcomes.

The medical records were abstracted by the first author (RM-S) and a research assistant who was trained by the first author. Neither abstractor was blinded to the study hypotheses. A 20 percent sample of the abstractions performed by the research assistant was recoded by the first author. The kappa statistic for interrater reliability for coding of diagnoses was .92 while the kappa for coding of treatments prescribed was .75. The majority of the discrepancies in coding between the abstractors were related to difficulty in interpreting the handwriting of the participating physicians.

Statistical Methods

Calculating the Proportion of Viral Cases Where Antibiotics Were Prescribed

The denominator for this measure was the number of cases in each sample of abstracted medical record encounters where *only* a viral diagnosis was made ($n = 91$ for 10/96–3/97 and $n = 77$ for 10/97–3/98). Viral diagnoses included all diagnoses listed under item 3 of the abstraction form that contain the word viral (see Appendix 1). In addition, all cases of bronchitis not otherwise specified (NOS) and croup were considered viral as well as cases of pharyngitis NOS and pneumonia NOS without laboratory values supporting a bacterial diagnosis (e.g., a positive rapid strep test, a CBC with a white blood cell [WBC] count $> 15,000$). The numerator for this measure was the number of cases in the denominator where an antibiotic was either provided or prescribed.

Calculating the Proportion of Cases Assigned a Bacterial Diagnosis

The denominator for this measure was the entire sample of abstracted medical record encounters for each data collection period ($n = 167$ for 10/96–3/97 and $n = 137$ for 10/97–3/98). The numerator was all encounters for each time period with a bacterial diagnosis. Bacterial diagnoses included all diagnoses listed under item 3 of the abstraction form that contain the word bacterial (see Appendix 1). In addition, all cases of acute otitis media, otitis externa, otitis media with effusion, pneumonia NOS (with a supporting chest x-ray or a CBC with a WBC count > 15,000), mycoplasma infection, culture or rapid antigen detection-test proven streptococcal pharyngitis, and sinusitis were considered bacterial.

Testing for Difference

Multiple logistic regression, correcting for clustering of encounters within physicians using the *cluster* option of the *logistic* command in STATA 6.0 (StataCorp. 1999), and controlling for practice site, and patient age was used to examine differences in the proportion of viral cases where antibiotics were prescribed between the observational and nonobservational studies. Multiple logistic regression, cor-recting for clustering of encounters within physicians was also used to examine differences in the proportion of cases assigned a bacterial diagnosis between the observational and nonobservational studies. Differences for which $p < .05$ were considered significant. All of the physicians (three) in one practice participated in the study, while in the second practice, all of the full-time clinicians (five) and two of five part-time clinicians, who work one-half day per week, participated in the study. Thus, study physicians account for 100 percent of encounters at one site and approximately 95 percent of encounters at the other site.

RESULTS

During the recruitment slide presentation given to the physicians prior to the observational study, one physician stated, "Parents in our practice probably won't be very happy since you are focusing on kids with colds and they all want antibiotics!" During the second slide presentation where the observational study results were reviewed, we asked the participating physicians if they had ascertained that we were most interested in how parent expectations affected their prescribing patterns. All of the physicians stated that they were not aware that this was our main objective. However, one physician did state

Figure 1:   Proportion of Viral Cases Prescribed Antibiotics Decreased When Physicians Were Directly Observed
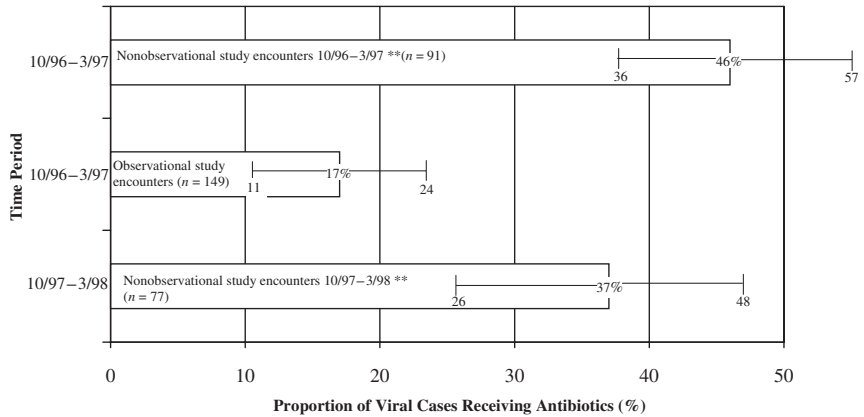
**p<.05 for comparisons of observational study period to both nonobservational study periods

that she thought we were probably interested in whether parents liked the way their doctor chose to treat their child.

Antibiotic prescribing rates and bacterial diagnosis rates did not differ significantly by month within provider for either the observational or nonobservational studies (data not shown).

For the observational study, the proportion of viral cases prescribed an antibiotic was 29 percentage points lower than the baseline proportion obtained from the nonobservational study during the time period 10/96 to 3/97 (Figure 1). These data represent unadjusted results correcting for clustering of encounters within physician only. We observed a decrease from 46 percent of nonobservational study viral cases receiving antibiotics to 17 percent of viral cases in the observational study ($p<.05$). During the following year (10/97–3/98), the unadjusted results indicated that antibiotics were prescribed for 37 percent of children diagnosed with viral illnesses ($p<.05$). These findings remained virtually the same after adjusting for practice site and patient age (Table 2).

Considering antibiotic prescribing for all encounters (including both bacterial and viral cases) during the period from 10/96 to 3/97, the proportion of visits where antibiotics were prescribed decreased from 68 percent (114/167) during the nonobservational study to 55 percent (147/272) during the observational study ($p<.05$).

Table 2:   Adjusted Predictors of Prescribing Antibiotics for a Presumed Viral Illness

Predictor Variable	Odds Ratio	95% CI for Odds Ratio
Encounter with no observation between 10/96 and 3/97*	3.60**	1.67–7.76
Encounter with no observation between 10/97 and 3/98*	2.61**	1.27–5.39
Practice site one	0.57	0.20–1.67
Patient age in years	0.95	0.70–1.30

*Omitted comparison group = Encounters that were observed during the observational study 10/96–3/97.
**$p < .01$

We also examined whether the proportion of cases assigned a bacterial diagnosis differed in the observational study compared to the nonobservational study. Overall, taking both study sites together, the proportion of cases assigned a bacterial diagnosis did not change between the observational study (45 percent of cases) and the first and second nonobservational study periods (45 percent and 44 percent of cases respectively; $p > 0.2$). However, when the sites were analyzed separately, the proportion of cases assigned bacterial diagnoses by physicians at one of the two sites significantly decreased during the observational study compared to the proportion of cases assigned bacterial diagnoses in the nonobservational study (Figure 2). During the observational study, participating physicians at this site assigned a bacterial diagnosis in 48 percent of the encounters. By comparison, nonobservational study visits abstracted from the period 10/96 to 3/97 for this site, were assigned a bacterial diagnosis in 77 percent of the encounters ($p < .05$). One year later, for visits abstracted from the period 10/97 to 3/98 for this site, physicians assigned a bacterial diagnosis in 68 percent of encounters ($p < .05$). For the second site, during the observational study, the proportion of cases assigned a bacterial diagnosis was 44 percent while for the two nonobservational study periods (10/96–3/97 and 10/97–3/98), the proportions of cases assigned bacterial diagnoses were 32 percent and 35 percent respectively ($p > 0.2$).

DISCUSSION

Although previous investigators have postulated that a perceived demand for improved performance or outcomes may be required for the Hawthorne effect to have a significant impact (Campbell, Maxey, and Watson 1995), the results of the current study indicate that this may not be the case. In this study,

Figure 2:   Physicians in One Practice Assigned Fewer Bacterial Diagnoses When under Direct Observation
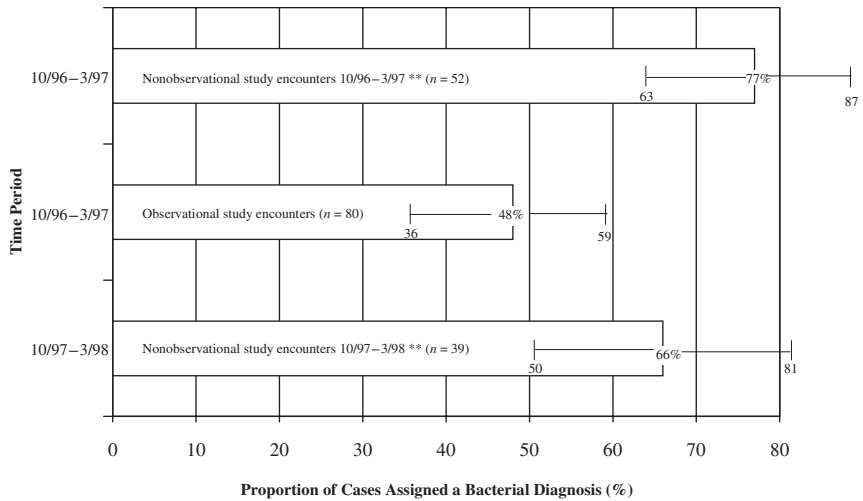
**p<.05 for comparisons of observational study period to both nonobservational study periods

physicians were not directly made aware that their antibiotic prescribing patterns and diagnostic decisions were the main outcomes of interest, yet they significantly altered their behavior with regard to these outcomes when they were surveyed and audiotaped. Although we attempted to blind physicians in the observational study with regard to our main outcomes of interest, that is, their antibiotic prescribing and diagnostic patterns, it is possible that they ascertained our intentions to study these particular outcomes. For example, the physicians might have been made aware of these objectives from the questions included on the physician postvisit survey (see Appendix 2). This instrument has a series of three items that inquire whether the physician perceived that the parent expected cough medicine, decongestants, or antibiotics to be prescribed during the visit. In addition, physicians were asked to indicate their diagnosis and what medications, if any, were prescribed. Just as blinding is frequently not possible for intervention trials that attempt to change physician practice patterns (Winkens et al. 1996) this also becomes an issue for purely observational studies. Although we could have assessed diagnostic decisions and prescribing from the medical record for the observational study encounters rather than directly surveying the physicians, we were only able to

ascertain physicians' perceptions of parent expectations by asking them directly. Thus our ability to completely blind them to our outcomes of interest was compromised. It seems likely that if the physicians in the observational study were even somewhat aware that we were examining their prescribing patterns, they might alter their management of the study cases to follow a "best practices" approach to treatment of their patients.

The proportion of viral cases in the nonobservational study prescribed antibiotics (46 percent and 37 percent) for the two data collection periods respectively (10/96–3/97 and 10/97–3/98) were remarkably similar to those obtained by Nyquist et al. in their two sequential studies that examined antibiotic prescribing rates for URIs based on data from the 1992 and 1996 National Ambulatory Medical Care Surveys (NAMCS). The values obtained in these surveys were 48 percent and 38 percent respectively (Nyquist et al. 1998b; 1998a). Although the time frames do not coincide exactly, the doctors in our sample were performing similarly to the national average in 1996 (38 percent) when they were not under direct observation in 1997–1998 (37 percent). Interestingly, the NAMCS data are collected by physicians completing standard surveys after a series of visits in their offices. Thus it is possible that the audiotaping procedures as well as the parental questionnaire in our observational study contributed to the changes in prescribing and diagnostic patterns that we observed. A separate analysis of the NAMCS data from 1997–1998 showed a similar secular trend toward decreased antibiotic prescribing for upper respiratory tract infections including acute otitis media, pharyngitis, bronchitis, and sinusitis (McCaig, Besser, and Hughes 2000). In this study 25 percent of such visits were prescribed antibiotics, which had decreased from 33 percent in 1989–1990. The percentage of visits where antibiotics were prescribed in 1997–1998 was markedly lower in this investigation than we have reported here. This is most likely explained by the fact that the two studies examined rates of prescribing for different respiratory conditions

Previous investigations have indicated that some inappropriate prescribing may be manifested as inappropriate diagnosis (Mangione-Smith et al. 1999; Vinson and Lutz 1993). In one of the two practices studied during the observational study, just as physicians lowered their frequency of prescribing antibiotics for viral illnesses, they also markedly decreased their frequency of making bacterial diagnoses. It could be argued that this reflected a true change in the frequency of observed bacterial illnesses between the observational study encounters and the nonobservational study encounters abstracted for 10/96–3/97, however, this seems unlikely since the time periods for data collection were the same. The only difference between the two types of encounters was

whether the child was in the observational study or not. The other practice included in this study, where physicians did not significantly alter how often they assigned bacterial diagnoses between the observational study and nonobservational study, is in the same locality as the practice where physicians changed their diagnostic patterns. Thus it does not seem plausible that the physicians changed their diagnostic patterns based on natural geographic variation in bacterial infectious diseases. Thus we believe the observed changes in bacterial diagnosis rates represent a true Hawthorne effect. The physicians who made fewer bacterial diagnoses during the observational study compared to the nonobservational study might have done so because they were again trying to take a "best practices" approach during the observational study encounters. Some of the bacterial diagnoses assigned by these same physicians during the non-observational study encounters may represent an attempt to justify giving an unnecessary antibiotic. This finding has implications for the development of future interventions to decrease inappropriate antibiotic prescribing rates. If the only outcome examined is antibiotic prescribing for viral illnesses, a substantial proportion of inappropriate prescribing will be missed in some practices. The current study suggests that changes from baseline rates of making bacterial diagnoses should also be measured.

The magnitude of the Hawthorne effect observed in the current study was similar to the percentage improvement observed in a recent intervention trial designed to decrease antibiotic prescribing for acute bronchitis in adult patients (29 percent versus 26 percent) (Gonzales et al. 1999). The investigators of this trial may have "harnessed" the Hawthorne effect as their intervention included supplying the participating physicians with feedback on their performance with regard to antibiotic prescribing for acute bronchitis in adults. Future research should focus on the sustainability of such improvements after this stimulus is removed. Is the knowledge that one's performance is being monitored necessary to maintain the observed improvements? Or will sustained improvement in performance continue without further monitoring?

The current study indicates that although we found an unacceptably high rate of inappropriate antibiotic prescribing in our observational study, particularly when physicians perceived parents as wanting antibiotics, this was an underestimate of the true magnitude of the problem.

Limitations

Because our study was done in one geographic location with a small and relatively homogeneous group of parents and physicians, we do not know whether our findings would generalize to other settings with parents from

different backgrounds. We may have introduced measurement error by using two different sources of data to determine diagnostic and antibiotic prescribing patterns for the two different studies: postvisit surveys of physicians for the observational study and medical record abstractions for the nonobservational study. It is possible that the diagnoses and treatment choices reported on the surveys did not match what was recorded in the medical record for the observational study visits. However, we believe the likelihood of this is small. We sampled one hundred of the encounters from the observational study in order to compare the physician's diagnosis recorded on the physician postvisit survey with the "official" diagnosis recorded in the medical record for the same encounter. For each physician the number of encounters sampled was proportional to the number of encounters contributed by that physician to the observational study. Diagnoses were found to be the same between the postvisit survey and the medical record in 93 percent of cases (95 percent confidence interval 86.1–97.1 percent). Although this finding does not rule out all bias, since the act of completing the physician survey may have influenced the medical record note, it does provide some support that the differences in diagnosis and treatment observed were not likely secondary to differences in the data collection methods used in the observational and nonobservational studies.

We did not review charts during the three two-week periods that the observational study occurred for each physician. Certain respiratory illnesses are more common in certain months of the year, for example streptococcal pharyngitis, which could thus change the diagnostic and treatment patterns for that month. Thus, because the observational and nonobservational cohort of visits for each physician in some cases represented different months during the first year (10/96–3/97) this might have affected the diagnosis and treatment rates obtained for each cohort. However, neither antibiotic prescribing rates nor bacterial diagnosis rates differed significantly by month within physician for either the observational or nonobservational studies. Thus, the discrepancies in calendar month for each physician between the observational and nonobservational study periods are unlikely to affect estimates of diagnostic or treatment behavior.

For the nonobservational study, the visits were "recruited" or screened based on a preselected list of diagnoses. At one practice this list was used to select cases from an administrative database. Once a claim with an appropriate diagnosis was identified, the chart with the eligible encounter was abstracted. At the second practice the list of diagnoses was used to screen visits for eligibility using manual review of visit notes in the medical record. These differences in case-finding strategies may have introduced bias into the study results. To measure the degree to which using these different strategies affected the results,

we performed a substudy at the practice site where administrative data were used to select charts for study inclusion. We randomly sampled one hundred charts and abstracted them for eligible diagnoses and treatments during 10/96–3/97 and 10/97–3/98 as was done at the other practice site during the nonobservational study. Among these one hundred abstracted encounters, the administrative data did not have a claim for the abstracted encounter 33 percent (19/57) of the time for encounters between 10/96 and 3/97, but for only 7 percent (3/43) of the abstracted encounters for 10/97–3/98. Thus the administrative data were incomplete in the first period during the observational study (10/96–3/97) but not the year after the observational study. The administrative database was first constructed in 1994–1995, which might explain the sharp decrease in the percentage of encounters without a claim in the database during 10/97–3/98. The fact that the conclusions are similar for both nonobservational study periods (see Figures 1 and 2) suggests that the missing data for 10/96–3/97 (while the data entry system was still undergoing refinement) might have been missing at random, and probably do not substantially bias the results. Of the 78 abstracted encounters that had claims in the administrative database, 97 percent had concordant diagnoses with the assigned ICD-9 codes in the database.

Neither medical record abstractor was blinded to the study hypotheses, which may have biased the results from the nonobservational study. Because of the simplicity of the abstraction form, this was felt to be a nonjudgement based clerical task and we do not believe the results were biased secondary to abstractor knowledge regarding the outcomes of interest.

We also do not know if physicians were truly unaware of our interest in antibiotic prescribing patterns for the observational study. Although we asked whether the physicians had ascertained our aims, we again were depending on physician self-reports for this information. It is possible that the physicians were aware that their prescribing behavior was under scrutiny and this led to their observed changes in behavior rather than just the awareness of being audio recorded and surveyed leading to these results.

CONCLUSIONS

The results of this study suggest two possible interpretations. First, the physicians may have altered both their diagnostic and prescribing patterns merely as a result of being studied. In this case, we would postulate that the participants were using a "best practices" approach to treating their patients with URIs because there was a tendency to improve performance in general

when being observed. Alternatively, the physicians may have ascertained our main outcomes of interest and thus felt the need to specifically improve on their performance with respect to prescribing antibiotics for children with URIs. In either case, this study offers some possible insights into the mechanisms that mediate the Hawthorne effect. In the case of antibiotic prescribing for URIs, the large Hawthorne effect may reflect the substantial role that nonmedical factors play in treatment decisions, and the minimal risk that accompanies inappropriate antibiotic treatment at the individual level during a particular encounter. Thus there may be a "flexibility" in prescription rates that can be susceptible to the Hawthorne effect. Contrast this with the role the Hawthorne effect might play in a study of febrile infants less than six weeks of age who are at risk for meningitis. Many parents are apprehensive to consent to a lumbar puncture because they fear their child will sustain damage to his or her spinal cord. However, in this case, parent expectations or demands might play a much less significant role because the risk of deviating from the standard of care would be much greater for the patient. In such cases, the Hawthorne effect could play a less prominent role in measurement error. It is probably most important to measure and adjust for the Hawthorne effect in observational studies that focus on treatment or management issues where the risk of deviating from the standard of care is minimal to the patient and where parent/patient expectations play a significant role. If such adjustments are not made, the results will likely underestimate the degree of the problem under study. Additionally, for intervention studies that do not specifically use profiling or feedback as components of the intervention, failure to make such adjustments will likely result in overestimates of the effectiveness of the intervention regarding performance improvement.

This study may also provide insights into the commonly observed deterioration in results from quality improvement studies. Physicians often revert to pre-intervention levels of performance after a study has concluded. Better understanding the mechanisms by which changes can be maintained is critical for the success of future quality improvement activities. Some ongoing stimulus may be needed to maintain improved performance.

ACKNOWLEDGMENTS

APPENDIX 1

Medical Record Abstraction Form   Study ID#: |_||_||_|

1. Date of Visit: |_||_|−|_||_|−|_||_|
                                           MO   DA   YR

2. Patient's Date of Birth: |_||_|−|_||_|−|_||_|
                                           MO   DA   YR

3. Record up to three assigned diagnosis: 1) |_|_| 2) |_||_| 3) |_||_|

Acute Otitis Media.…………….…………….…………01	Mycoplasma Infection……...……………. ….11
Otitis Media with Effusion (OME)……….………… 02	Croup…………………………………………..12
Otitis Externa.……………….…………….……... 03	Streptococcal Pharyngitis………………… ….13
Asthma.…………….…………….…………….04	Viral Pharyngitis.……………………………14
Bacterial Bronchitis.…………….……………. 05	Pharyngitis NOS.…………………………….15
Viral Bronchitis.…………….…………….……06	Viral Upper Respiratory Infection………….…16
Bronchitis NOS.…………….…………….... 07	Viral Stomatitis.…………….………………17
Bacterial Pneumonia.…………….…………..…08	Sinusitis.…………….…………….……...….18
Viral Pneumonia.…………….…………….......09	Viral Syndrome.…………...………………….19
Pneumonia NOS.…………….…………….....10	Other.………………………………………….20

4. Was a rapid strep test performed? |_|

       Yes………….1
       No…….……2

       If yes, was it positive? |_|

       Yes………….1
       No…….……..2

5. Was a throat culture obtained? |_|

       Yes………….1
       No…….……2

       If yes, was it positive? |_|

       Yes………….1
       No…….……2

6. Was a chest x-ray performed? |_|

       Yes………….1
       No…….……2

       If yes, was it positive? |_|
       Yes………….1
       No…….……..2

7. Was a CBC performed? |_|

       Yes……….…1
       No…….……...2

If yes, WBC count result: |_|

<10,000……………....1
10,000 – 15,000…..…..2
>15,000……………....3

8. Was a sinus x-ray or CT |_|
   performed?

       Yes……….…1
       No…….……..2

If yes, was it positive? |_|

       Yes………….1
       No…..……….2

9. Were oral antibiotics prescribed/provided? |_|

       Yes……………1
       No……………..2

If yes, what antibiotic was prescribed? |_||_|

Amoxicillin…………………………….…..………01
Amoxicillin and Clavulanic Acid (Augmentin)….…02
Azithromycin (Zithromax)…………………………03
Cefaclor (Ceclor)………………………………….04
Cefixime (Suprax)………………………………….05
Cefpodoxime(Vantin)………………………………06
Cefprozil(Cefzil)…………………………………..07
Cetriaxone (Rocephin)……………………………08
Clarithromycin (Biaxin)…………………………..09
Erythromycin and Sulfisoxazole (Pediazole)……..10
Loracarbef (Lorabid)…………………………… 11
Trimethoprim and Sulfamethoxazole
(Bactrim, Septra) ………………………………...12
Other………………………………………………13
No Data.……………………………………....…99

APPENDIX 2

Study ID#: _____   **MD Code#:**_____

MD Post - Visit Checklist

1) Please complete this checklist after each visit with a study parent and child.
Unless otherwise noted, please <u>check</u> all categories that apply to this visit.

General Physical Findings : (<u>Check</u> all that apply)

____Cervical Lymphadenopathy
____Fever
____Oral Vesicles
____Pharyngeal Erythema
____Pharyngeal Exudates
____Rhinorrhea

Tympanic Membrane (<u>Check</u> all that apply)

<u>Left</u>		<u>Right</u>
____	Bulging	____
____	Draining Pus	____
____	Fluid Behind TM	____
____	Mobility, Good	____
____	Mobility, Poor	____
____	Mobility, not evaluated	____
____	Opaque/Dull	____
____	Perforated	____
____	Pressure Equalizing Tubes	____
____	Red	____
____	Retracted	____
____	Not Visualized	____
____	Normal	____

Diagnosis : (<u>Check</u> all that apply)

____Acute otitis media	____Herpangina	____ Viral Pharyngitis
____Allergic Rhinitis	____Otitis Externa	____Viral Stomatitis
___Bacterial Pharyngitis	____Otitis Media with Effusion	____Viral URI
____Croup	____Sinusitis	____Bronchitis
		Other: _____
		(please specify)

Treatment: (<u>Check</u> all that apply)

____Analgesic/Antipyretic   ____Combination Antihistamine/
                                                                              Decongestant
____Antihistamine   ____Cough Medicine

____Antibiotic   ____Decongestant
  ____Amoxicillin
  ____Amoxicillin and Clavulanic Acid (Augmentin)
  ____Azithromycin (Zithromax)   ____Other: _____
  ____Cefaclor (Ceclor)   (please specify)
  ____Cefixime (Suprax)
  ____Cefpodoxime (Vantin)
  ____Cefprozil (Cefzil)
  ____Clarithromycin (Biaxin)
  ____Erythromycin and Sulfisoxazole (Pediazole)
  ____Trimethoprim and Sulfamethoxazole (Bactrim, Septra)

2) We would like to know your perceptions about some aspects of the visit you just completed.
Please <u>circle</u> <u>the</u> <u>number</u> that most closely reflects your assessment of the visit.

<u>CIRCLE ONE NUMBER ON EACH LINE</u>

	Strongly <u>Agree</u>	Somewhat <u>Agree</u>	<u>Uncertain</u>	Somewhat <u>Disagree</u>	Strongly <u>Disagree</u>
a) The cause of this child's illness was viral...............1	2	3	4	5
b) This parent expected me to recommend cough medicine………………………………1	2	3	4	5
c) I am very certain about the diagnosis in this case……………………………………1	2	3	4	5
d) This parent expected me to prescribe antibiotics…………………………………………1	2	3	4	5
e) The cause of this child's illness was bacterial……………………………………………1	2	3	4	5
f) This parent did not ask for a particular treatment…………………………………1	2	3	4	5
g) This parent was satisfied with the visit………………………………………….…1	2	3	4	5
h) The cause of this child's illness was both viral and bacterial ..………………………..1	2	3	4	5

**THANK YOU VERY MUCH FOR COMPLETING THIS QUESTIONNAIRE.
PLEASE PLACE IT IN THE BOX MARKED "MD SURVEYS" WHEN YOU ARE DONE.
IF YOU HAVE ANY QUESTIONS, PLEASE CONTACT RITA MANGIONE-SMITH AT
(310) 825-1988**

REFERENCES

Arborelius, E., and T. Timka. 1990. "In What Way May Videotapes Be Used to Get Significant Information about Patient–Physician Relationship?" *Medical Teacher* 12 (2): 197–208.

Barden, L. S., S. F. Dowell, B. Schwartz, and C. Lackey. 1998. "Current Attitudes Regarding Use of Antimicrobial Agents: Results from Physicians' and Parents' Focus Group Discussions." *Clinical Pediatrics* 37 (11): 665–72.

Butler, C. C., S. Rollnick, R. Pill, F. Maggs-Rapport, and N. Stott. 1998. "Understanding the Culture of Prescribing: Qualitative Study of General Practitioners' and Patients' Perceptions of Antibiotics for Sore Throats." *British Medical Journal* 317 (7159): 637–42.

Campbell, J. P., V. A. Maxey, and W. A. Watson. 1995. "Hawthorne Effect: Implications for Prehospital Research." *Annals of Emergency Medicine* 26 (5): 590–4.

Carabin, H., T. W. Gyorkos, J. C. Soto, L. Joseph, P. Payment, and J.-P. Collet. 1999. "Effectiveness of a Training Program in Reducing Infections in Toddlers Attending Day Care Centers." *Epidemiology* 10 (3): 219–27.

Davis, D. A., M. A. Thompson, A. D. Oxman, and B. Haynes. 1995. "Changing Physician Performance, a Systematic Review of the Effect of Continuing Medical Education Strategies." *Journal of the American Medical Association* 274 (9): 700–5.

Davy, T., P. T. Dick, and P. Munk. 1998. "Self-reported Prescribing of Antibiotics for Children with Undifferentiated Acute Respiratory Tract Infections with Cough." *Pediatric Infectious Disease Journal* 17 (6): 457–62.

De Amici, D., C. Klersy, F. Ramajoli, L. Brustua, and P. Politi. 2000. "Impact of the Hawthorne Effect in a Longitudinal Clinical Study: The Case of Anesthesia." *Controlled Clinical Trials* 21 (2): 103–14.

Dosh, S. A., J. M. Hickner, A. G. Mainous, and M. H. Ebell. 2000. "Predictors of Antibiotic Prescribing for Nonspecific Upper Respiratory Infections, Acute Bronchitis, and Acute Sinusitis." *Journal of Family Practice* 49 (5): 407–14.

Gonzales, R., J. F. Steiner, A. Lum, and P. H. Barrett. 1999. "Decreasing Antibiotic Use in Ambulatory Practice. Impact of a Multidimensional Intervention on the Treatment of Uncomplicated Acute Bronchitis in Adults." *Journal of the American Medical Association* 281 (16): 1512–9.

Greco, P. J., and J. M. Eisenberg. 1993. "Changing Physician Practices." *New England Journal of Medicine* 329 (17): 1271–3.

Grufferman, S. 1999. "Complexity and the Hawthorne Effect in Community Trials." *Epidemiology* 10 (3): 209–10.

Le Saux, N., B. Pham, C. Bjornson, and C. Pitters. 1999. "Antimicrobial Use in Febrile Children Diagnosed with Respiratory Tract Illness in an Emergency Department." *Pediatric Infectious Disease Journal* 18 (12): 1078–80.

Lohr, K. N., R. H. Brook, and M. A. Kaufman. 1980. "Quality of Care in the New Mexico Medicaid Program (1971–1975), the Effect of the New Mexico Experimental Medical Care Review Organization on the Use of Antibiotics for Common Infectious Diseases." *Medical Care* 18 (1): 1–129.

Mainous, A. G., W. J. Hueston, and M. M. Love. 1998. "Antibiotics for Colds in Children. Who Are the High Prescribers?" *Archives of Pediatrics and Adolescent Medicine* 152 (4): 349–52.

Mangione-Smith, R., E. A. McGlynn, M. N. Elliott, P. Krogstad, and R. H. Brook. 1999. "The Relationship between Perceived Parental Expectations and Pediatrician Antimicrobial Prescribing Behavior." *Pediatrics* 103 (4): 711–8.

McCraig, L., R. Besser, and J. Hughes. 2000. "Decline in Pediatric Antimicrobial Drug Prescribing Among Office-Based Physicians in the United States, 1989–1998." Proceedings of the Infectious Diseases Society of America Annual Meeting, September 7–10, 2000, New Orleans, LA. Abstract.

Nyquist, A., R. Gonzales, J. F. Steiner, and M. A. Sande. 1998a. "Antibiotics for Children with Upper Respiratory Infections." *Journal of the American Association* 280 (16): 1401. Letter.

———. 1998b. "Antibiotic Prescribing for Children with Colds, Upper Respiratory Tract Infections, and Bronchitis." *Journal of the American Medical Association* 279 (11): 875–7.

Roethlisberger, F. H., and W. J. Dickson. 1939. *Management and the Worker.* Cambridge, MA: Harvard University Press.

Schwartz, R. H., B. J. Freij, M. Ziai, and M. J. Sheridan. 1997. "Antimicrobial Prescribing for Acute Purulent Rhinitis in Children: A Survey of Pediatricians and Family Practitioners." *Pediatric Infectious Disease Journal* 16 (2): 185–90.

StataCorp. 1999. *Stata Statistical Software*, Release 6.0. College Station, TX: Stata Corportation.

Vinson, D. C., and L. J. Lutz. 1993. "The Effect of Parental Expectations on Treatment of Children with a Cough: A Report from the ASPN." *Journal of Family Practice* 37 (1): 23–7.

Watson, R. L., S. F. Dowell, M. Jayaraman, H. Keyserling, M. Kolczak, and B. Schwartz. 1999. "Antimicrobial Use for Pediatric Upper Respiratory Infections: Reported Practice, Actual Practice, and Parent Beliefs." *Pediatrics* 104 (6): 1251–7.

Winkens, R., J. Knottnerus, A. Kester, R. Grol, and P. Pop. 1996. "Fitting a Routine Health-Care Activity into a Randomized Trial: An Experiment Possible without Informed Consent?" *Journal of Clinical Epidemiology* 50 (4): 435–9.