The Effect of Expert Feedback on Antibiotic Prescribing in Pediatrics: Experimental Evidence

Kerstin Eilermann, Katrin Halstenberg, Ludwig Kuntz, Kyriakos Martakis,
Bernhard Roth, and Daniel Wiesen

Background. Inappropriate prescribing of antibiotics, which is common in pediatric care, is a key driver of antimicrobial resistance. To mitigate the development of resistance, antibiotic stewardship programs often suggest the inclusion of feedback targeted at individual providers. Empirically, however, it is not well understood how feedback affects individual physicians' antibiotic prescribing decisions. Also, the question of how physicians' characteristics, such as clinical experience, relate to antibiotic prescribing decisions and to responses to feedback is largely unexplored. **Objective.** To analyze the causal effect of descriptive expert feedback (and individual characteristics) on physicians' antibiotic prescribing decisions in pediatrics. **Design.** We employed a randomized, controlled framed field experiment, in which German pediatricians ($n = 73$) decided on the length of first-line antibiotic treatment for routine pediatric cases. In the intervention group ($n = 39$), pediatricians received descriptive feedback in form of an expert benchmark, which allowed them to compare their own prescribing decisions with expert recommendations. The recommendations were elicited in a survey of pediatric department directors ($n = 20$), who stated the length of antibiotic therapies they would choose for the routine cases. Pediatricians' characteristics were elicited in a comprehensive questionnaire. **Results.** Providing pediatricians with expert feedback significantly reduced the length of antibiotic therapies by 10% on average. Also, the deviation of pediatricians' decisions from experts' recommendations significantly decreased. Antibiotic therapy decisions were significantly related to pediatricians' clinical experience, risk attitudes, and personality traits. The effect of feedback was significantly associated with physicians' experience. **Conclusion.** Our results indicate that descriptive expert feedback can be an effective means to guide pediatricians, especially those who are inexperienced, toward more appropriate antibiotic prescribing. Therefore, it seems to be suitable for inclusion in antibiotic stewardship programs.

Keywords

clinical experience, descriptive feedback, expert benchmark, framed field experiment, length of antibiotic therapy

Inappropriate use of antibiotics is widespread and contributes to rapidly increasing antimicrobial resistance (AMR),[1,2] which has become a serious global public health problem.[3,4] Besides the choice of the antibiotic agent, the dosage, and the correct initiation, the length of therapy is relevant for an appropriate antibiotic treatment.[5,6] Excessive use of antibiotics and unnecessarily long treatment courses have a significant impact on the development of AMR.[1,7] In pediatrics, inappropriate antibiotic prescribing is a particular concern due to antibiotic-related adverse outcomes, such as organ toxicity.[8–10] The need for effective measures to support physicians practicing in

Corresponding Author

Daniel Wiesen, Department of Business Administration and Health Care Management, University of Cologne, Albertus-Magnus-Platz, Cologne, 50923, Germany (wiesen@wiso.uni-koeln.de).

pediatric and neonatal settings is therefore an urgent issue.[11] Antibiotic stewardship programs pick up on this and often suggest the inclusion of feedback mechanisms targeted at antibiotic prescribing practices of individual providers.[1,12,13]

Empirically, however, it is not well understood how feedback causally affects *individual* physicians' antibiotic prescribing and whether their characteristics, such as clinical experience, relate to their responses to feedback. Some studies using cross-sectional data report that feedback can be effective in achieving more appropriate antibiotic prescribing.[14–16] Nevertheless, cross-sectional data may suffer from multiple confounding effects (e.g., lack of control, self-selection, or simultaneous policy interventions and institutional changes), making causal inferences difficult.[17,18] Furthermore, evidence from randomized controlled experiments on the effectiveness of feedback in medical practice is rather mixed.[19] Systematic evidence relating to antibiotic prescribing is scarce.[13] A recent randomized controlled trial (RCT) in the United Kingdom reported that providing social norm feedback affects general practitioners' antibiotic prescribing behavior.[20] Similarly, an RCT in the United States found that peer comparison among primary care practitioners decreases overall antibiotic prescribing rates at the practice level.[21] However, the causal effect of feedback on antibiotic prescribing at the level of *individual* physicians remains barely understood.

The main objective of our study was to analyze the causal effect of expert feedback, a descriptive norm, on antibiotics prescribing in pediatrics. We considered physicians' individual prescribing decisions in a randomized, controlled, framed field experiment with 73

Cologne Graduate School in Management, Economics, and Social Sciences (CGS), Department of Business Administration and Health Care Management, University of Cologne, Cologne, Germany (KE); Medical Faculty and University Hospital, Department of Pediatrics, University of Cologne, Cologne, Germany (KH, KM, BR); Department of Business Administration and Health Care Management, University of Cologne, Cologne, Germany (LK, DW); Operations Management Group, Judge Business School, University of Cambridge, Cambridge, UK (LK); Department of International Health, Care and Public Health Research Institute, School CAPHRI, Maastricht University, Maastricht, the Netherlands (KM); and Department of Pediatric Neurology, University Children's Hospital (UKGM) and Medical Faculty, Justus Liebig University of Giessen, Giessen, Germany (KM).

pediatricians.[a] In our experiment, which followed a mixed factorial design, pediatricians decided on the length of antibiotic therapies for hypothetical routine cases of pediatric infectious diseases. In the intervention group, we first announced that feedback would be given and then provided expert feedback. Pediatricians received an aggregate expert recommendation (expert benchmark) on the appropriate length of therapies to which they could compare their own (aggregated) decisions. The control group did not receive any feedback. The expert recommendations were elicited in a survey of directors of pediatric departments in Germany ($n = 20$).

Our study relates to recent social-norm feedback interventions in health care.[20,21] In our feedback mechanism, we employed an expert benchmark as a descriptive norm to guide pediatricians toward appropriate antibiotic prescribing. Our aim was to avoid potential adverse effects of comparisons with peers, such as an unintended change in the behavior of those performing better than the average toward the peers' average (the so-called "boomerang-effect").[25–27] The psychological literature on social norms provides evidence that descriptive normative information is an effective tool for changing behavior and for reducing undesired conduct.[28,29] We thus hypothesized that giving pediatricians expert feedback, which conveys a descriptive norm for antibiotic prescribing, would affect decisions on the length of antibiotic therapies and increase the appropriateness of prescribing.

We focused on the *length* of antibiotic therapy as it is critical for outcomes in children and for the development of antibiotic resistance.[5,6] Despite its importance, the length of antibiotic therapy has been neglected in studies on feedback interventions aimed at improving antibiotic prescribing. Existing studies rather focus on the choice of antibiotic agents or whether antibiotic therapies are initiated or not.[16,20,21] We thus complement this literature by providing evidence on the causal effect of feedback on the length of antibiotic therapies.
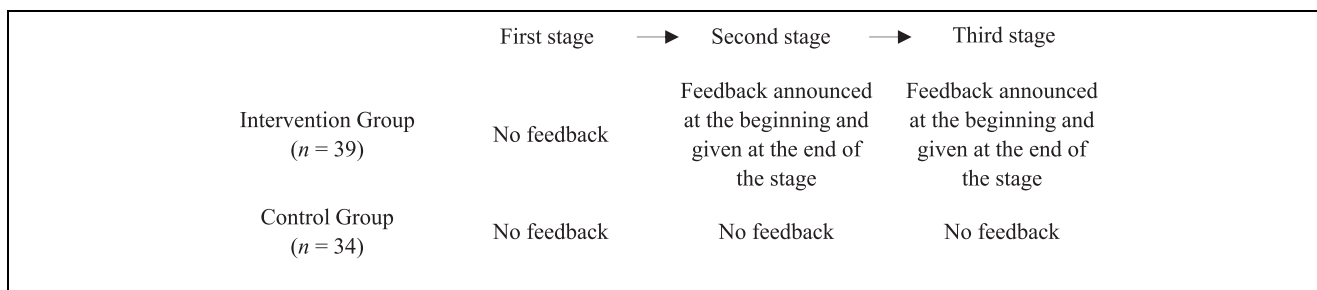
Furthermore, we investigated whether and how pediatricians' individual characteristics, including sex, clinical experience, risk attitudes, and personality traits, relate to antibiotic therapy decisions. We thus contribute to a recent stream of literature linking physicians' characteristics to medical treatment decisions. Current evidence suggests

---

[a]According to Harrison and List's widely used taxonomy of behavioral experiments, which ranges from laboratory experiments to natural field experiments, a framed field experiment is a structured experiment with subjects making decisions in their natural environment with the familiar context of the task, stakes, or information set.[22–24]

	First stage	→	Second stage	→	Third stage
Intervention Group (*n* = 39)	No feedback		Feedback announced at the beginning and given at the end of the stage		Feedback announced at the beginning and given at the end of the stage
Control Group (*n* = 34)	No feedback		No feedback		No feedback

Figure 1 Stages of the experiment. In each stage of the experiment, subjects decided on the length of antibiotic therapy for 40 routine cases, which were shown in randomized order. The first stage was the same in the intervention and control groups. At the beginning of the second stage, the intervention group was told that feedback would be given. After the second stage, feedback was shown such that subjects could compare the average of their chosen length of antibiotic therapies with the expert benchmark. The third stage was analogous to the second stage. In the control group, the decision situations in the second and third stages were identical to those in the first stage, and no feedback was announced or given.

that medical service provision is related to physicians' risk attitudes[30–35] and experience.[36,37] Furthermore, the sex of physicians is associated with treatment[36] and prescribing decisions[38] and with patient outcomes.[39,40] Personality traits are also important to explain decisions and behavior in various contexts.[41,42] A few recent studies aim to link personality traits to the behavior of health care providers.[43,44] While these characteristics seem to be relevant in explaining the behavior of physicians, their association with antibiotic prescribing decisions remains inconclusive.[45,46] With respect to antibiotic prescribing, only the role of experience has been studied to a somewhat larger extent. Evidence from primary care settings suggests a positive association between physicians' years of experience and their willingness to prescribe antibiotics.[47–49]

To contribute to a better understanding of how provider characteristics affect antibiotic prescribing, we related pediatricians' decisions on the length of antibiotic therapies to their sex, clinical experience, risk attitudes, and personality traits. In a comprehensive postexperimental questionnaire, we elicited the pediatricians' demographics, personality traits (using the Big Five inventory[50,51]), and social and risk preferences.[52–54] We linked information on the pediatricians' characteristics to their decisions made in the experiment and controlled for the potential impact of characteristics in our regression analyses.

We further analyzed how responses to expert feedback are related to clinical experience. Drawing on the theory of knowledge[55,56] and theories of learning and routines,[57–59] which imply that humans develop knowledge, specific capabilities, and routines mainly through repetition and experiential hands-on learning, we hypothesized that physicians with more experience would be less prone to adapt their decisions after receiving expert feedback but would rather tend to follow their own routines (built, for example, through hands-on experience with patients). We assumed that less experienced physicians would rely more on external input and hence be more likely to adapt their decisions.

In sum, our study addressed the main research question of how expert feedback causally affects individual pediatricians' decisions on (i) the length of antibiotic therapies and (ii) the appropriateness of antibiotic therapy decisions. We also investigated (iii) how pediatricians' individual characteristics relate to antibiotic prescribing decisions and (iv) how pediatricians' clinical experience relates to responses to expert feedback.

Methods

The Experiment: Design

Our framed field experiment comprised three stages. In each stage, pediatricians decided on the length of first-line antibiotic therapies for 40 routine pediatric cases, which were shown on the subjects' computer screens in randomized order. For each case, the pediatricians decided on the length of antibiotic therapy by entering an integer between 0 and 28 in an open field below the respective case description. In total, each pediatrician made 120 decisions in the three stages of the experiment. For completing the task, participants received a lump-sum payment of €50.

Pediatricians were randomly allocated to either an intervention or a control group. In the intervention group, we introduced feedback in form of an expert benchmark at the within-subject level (see Figure 1). In the first stage, no feedback was provided. In the second

stage, we announced that feedback would be provided at the end of the stage. After the second stage, feedback was shown (graphically as bar charts and numerically) such that subjects could compare their average length of antibiotic therapies for the 40 cases with the expert benchmark; for a sample screen, see Figure A.1 in Online Appendix A.1. The third stage was analogous to the second stage. This design allowed us to disentangle the effect of announcing feedback (comparing decisions from the first and second stages) and the effect of providing feedback (comparing decisions from the second and third stages). In the control group, feedback was neither announced nor provided in any stage. For the instructions of the experiment, see Online Appendix A.2.

Medical Cases and Expert Benchmark

The 40 cases covered a broad range of typical infectious diseases in pediatrics, namely, i) neonatal infections, ii) infections of the central nervous system, iii) bone and joint infections, iv) upper respiratory tract infections, v) lower respiratory tract infections, and vi) urinary tract infections; Online Appendix A.3.1 provides the case descriptions. The case scenarios had been developed by the clinicians in the research team (three pediatricians with different subspecializations) based on their clinical experience, clinical case reports, and textbooks. Afterward, the cases were validated by five pediatricians of the Department of Pediatrics at the University Hospital Cologne, who did not participate in the experiment. The aim was to ensure i) clarity and comprehensibility of the cases, ii) their relevance in clinical practice, iii) their plausibility, and iv) correctness and completeness of the given information; for more details, see Online Appendix A.3.2.

For all cases, the study participants decided on the length of first-line antibiotic therapy, which could be between 0 and 28 days. Besides the length of therapy, the class of antibiotics and the dosage play a role for treatment outcomes.[60] We asked the pediatricians to consider the standard antibiotic agent and the standard dosage for each case when deciding on the length of therapies. We designed the cases such that a standard antibiotic agent and a standard dosage were available for all cases for which antibiotic treatment was indicated; see Online Appendix A.3.2 for details. We did not specify the agent to be used, as we intended to leave the decision on whether any antibiotics should be prescribed to the discretion of the pediatrician. With the option to choose 0 days of antibiotic therapy, the task includes the decision on whether to initiate antibiotic therapy or not.

By using an expert benchmark as a norm for antibiotic prescribing, we contribute to the literature on the use of benchmarks in health care.[61] The expert benchmark is a descriptive norm because it provides information on the decisions of others for purposes of comparison.[28,29] We chose experts to define a normative benchmark that reflects personal expertise, national medical guidelines, and local standards in pediatric departments. To form the benchmark, we surveyed directors of German pediatric departments (referred to as "experts") on their recommended length of antibiotic therapies for the 40 cases we used in the experiment. In total, 50 randomly chosen directors were contacted by formal letter, in which we asked them about their willingness to participate in a survey; 20 directors participated in our online survey between September and October 2014. As the expert benchmark, we chose the length of therapy averaged over all cases and experts, which was 6.42 days (SD, 4.94; 95% confidence interval [CI], 4.26 to 8.59).

To qualify the experts' decisions and to assess their suitability for a normative benchmark, we compared them with published recommendations on the length of antibiotic therapies. In particular, we considered recommendations published by the German Society for Pediatric Infectious Diseases for comparison.[62] While we observed some variation, the experts' decisions imply a high compliance with the recommendations; see Online Appendix A.4 for details. Besides the fact that the experts' recommendations were close to guidelines, we chose the aggregated expert benchmark as a means of feedback: i) to maintain the pediatricians' discretion in choosing the cases for which, if at all, they would change their initially chosen length of therapy; ii) to mimic a simple feedback mechanism that could potentially be implemented in a real clinical setting, as providing feedback on a case-by-case basis seems prohibitively challenging; and iii) to provide pediatricians in the experiment with a simple *directional* reference that could guide their own decisions.

Providing pediatricians with an expert benchmark that allows comparing one's own decisions with an expert recommendation is distinct from feedback applying peer comparisons (e.g., relative performance compared to peers). The latter may have unintended effects such as the previously mentioned "boomerang effect";[25,26] see Linder[27] on the importance of the design of feedback mechanisms and Meeker et al.,[21] who used a similar approach by allowing for comparisons with top performers instead of average-performing peers in their feedback intervention.

We employed a benchmark based on the opinion of experts instead of guideline recommendations, because physicians' negative attitudes toward medical guidelines have been identified as one of the main reasons for low

guideline compliance in clinical practice.[63] Major concerns include the flexibility and applicability of guidelines in general and, in particular, antibiotic treatment recommendations.[63,64] Qualitative research has shown that other clinicians' opinions are the main source of knowledge about antibiotic prescribing in clinical practice. The opinions of other medical professionals have a greater impact on antibiotic prescribing decisions and are perceived as more effective in modifying prescribing patterns than guideline recommendations.[65] Based on these findings, we assumed expert-based feedback, reflecting the opinions of German pediatric department directors, to have a potentially greater effect than guideline-based feedback.

Sample and Procedure

The computerized experiment was conducted with mobile tablet computers of the Cologne Laboratory for Economic Research (CLER). The experiment was programmed in z-Tree.[66] Experimental sessions took place at the Department of Pediatrics at the University Hospital Cologne (October and December 2014), the Children's Hospital of the City of Cologne (June 2015), and during the annual conference for pediatricians (Päd-Ass 2015) in Cologne (March 2015). The experiments were conducted in hospital seminar rooms, which we equipped with tablet computers and cubicles to ensure anonymous decision making; for an illustration, see Figure A.2 in Online Appendix A.5.

Sample size calculations showed that at least 32 subjects in each experimental group were necessary to detect a difference of 0.5 days between the 2 groups, considering changes from stage 2 to stage 3 in both groups (between-subject comparison), using a two-tailed Mann-Whitney U test, and assuming a power of 80% and a 5% significance level. Prestudy sample size calculations were conducted using G*Power[67]; for more details, see Online Appendix A.6.

Overall, 73 pediatricians participated in our experiment; directors of pediatric departments were excluded. Pediatricians were recruited via e-mail and posters, which provided general information about the experiment and the scheduled sessions. Pediatricians were allowed to register only for one of the sessions publicized through an online poll. In total, 8 sessions were conducted. In sessions at the Department of Pediatrics at the University Hospital Cologne and the Päd-Ass conference 2015, 22 and 6 subjects participated in the intervention and 14 and 20 in the control groups, respectively. At the Children's Hospital of the City of Cologne, 11 subjects participated in the intervention group.

Table 1 Baseline Characteristics of the Study Population

	Intervention Group ($n = 39$)	Control Group ($n = 34$)
Sex, n (%)		
Male	11 (28)	6 (18)
Female	28 (72)	28 (82)
Share of consultants, n (%)	15 (39)	12 (35)
Experience (years worked in hospital), mean (SD)	5.37 (4.66)	5.05 (5.98)

Using a simple coin toss, it was randomly determined whether intervention (feedback) or control treatment would be employed in a particular session. Pediatricians, uninformed about the content of the experiment prior to participation, were therefore allocated randomly to one of the two experimental groups. The baseline characteristics of the participants were well balanced between the two groups; see Table 1.

We employed a double-blind procedure. The person who conducted the experiment and managed the data was not involved in the recruiting of subjects. For each session, an external research assistant, employed by the Department of Personnel Economics of the University of Cologne, facilitated subject recruitment, registration, and remuneration. Upon their arrival, pediatricians drew a number that indicated their cubicle and computer. Decisions on the computer screens were made anonymously; the experimenter was only able to link the randomly assigned computer number to the respective subject's data. Payment was handed out in sealed envelopes.

The experimental sessions lasted for about 1 hour. Before the experiment started, written informed consent was obtained from all subjects, and they received written instructions describing the general structure, the decision situation, and the task of the experiment. Prior to each stage of the experiment, subjects received stage-specific instructions. They were given sufficient time to read the instructions, and any upcoming questions were answered in private at the cubicles. After completing the experiment and before receiving their payment, subjects were asked to answer some questions on their demographics and practical experience. Furthermore, we elicited subjects' personality traits using the short 10-item Big Five questionnaire[50,51] and their economic preferences, including risk attitudes, using validated survey questions[52–54]; for the full questionnaire, see Online Appendix A.7. One month after the study had been concluded, debriefings

with participating pediatricians and heads of pediatric clinics took place.[68]

Statistical Analyses

To determine the effect of expert feedback on the length of antibiotic therapies and on the appropriateness of the length of therapies, we employed nonparametric statistical analyses. At the within-subject level, we compared the length of therapies and the absolute deviation from the expert recommendations between the three stages in both experimental groups. We assessed the impact of merely announcing feedback (differences between the first and second stages) and of actually providing feedback (differences between the second and third stages), using 2-sided Fisher-Pitman permutation tests for paired replicates. For between-subject comparisons, we used 2-sided Fisher-Pitman permutation tests for independent samples. We also employed Mann-Whitney U and Wilcoxon signed-rank tests for between-subject and within-subject comparisons, respectively.

To account for heterogeneity in the experimental data, we ran a series of multilevel mixed-effects panel regression models. For details on the model specification, see Online Appendix B.

To analyze the association between pediatricians' individual characteristics and their antibiotic therapy decisions, we employed multilevel mixed-effects models. For this analysis, we only considered the decisions from the first stage of the experiment when the instructions were the same for subjects in the control and in the intervention group. The statistics software STATA 14.1 (StataCorp, College Station, TX) was used for all analyses.

Role of the Funding Source

The funding source had no role in the study design or implementation.

Results

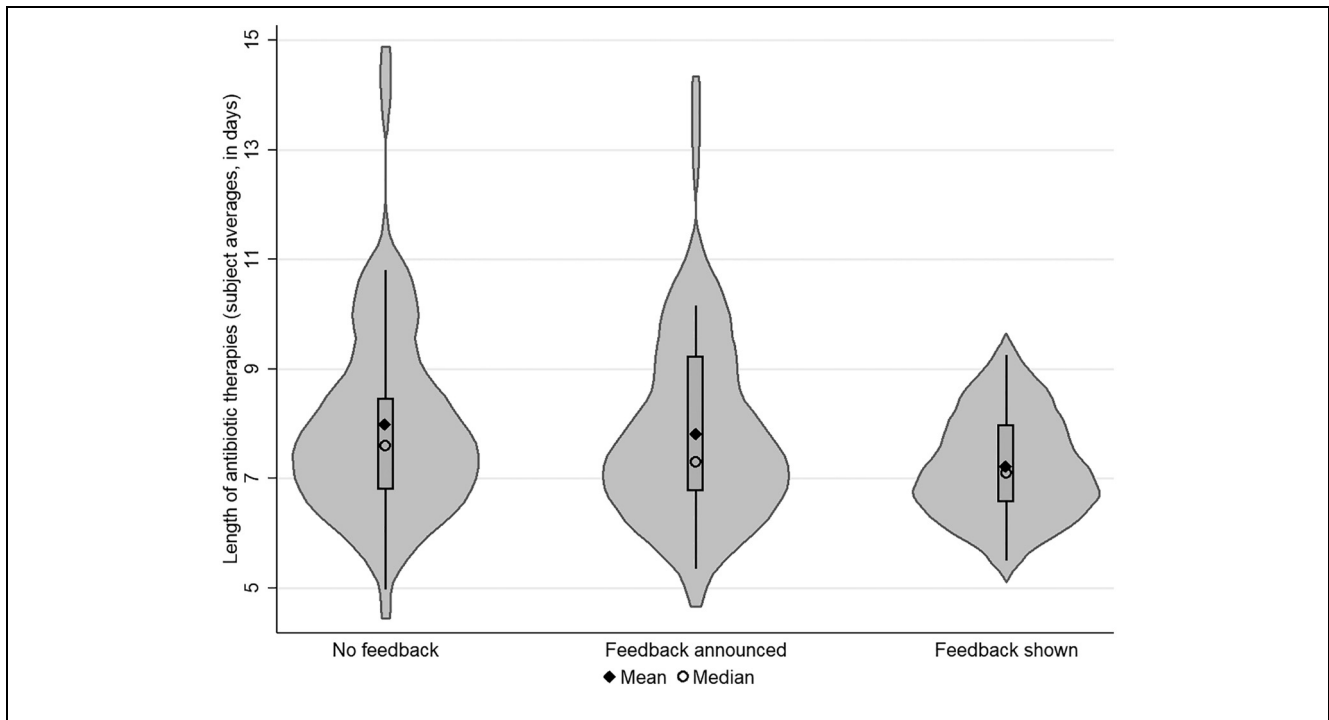### The Effect of Feedback on Antibiotic Prescribing

First, we analyzed the effect of feedback on pediatricians' decisions at a within-subject level in both groups. In the intervention group, the average length of antibiotic therapy was 7.98 days (95% CI, 7.42 to 8.53, $n = 1,560$) in the first stage. After the announcement of feedback (in the second stage), the average number of days fell slightly to 7.83 (95% CI, 7.31 to 8.35; $n = 1,560$), which was not statistically significant ($P = 0.153$, Fisher-Pitman

permutation test for paired replicates). In the third stage, when pediatricians had compared their average length of antibiotic therapies (from the second stage) with the expert benchmark, the mean length of antibiotic therapies fell to 7.23 days (95% CI, 6.93 to 7.53; $n = 1,560$). Providing pediatricians with the expert benchmark significantly reduced the length of antibiotic therapies ($P = 0.000$, Fisher-Pitman permutation test for paired replicates). For an illustration of how the decisions in the intervention group changed between the stages, see Figure 2. Changes between the stages in the control group were not significant (both $P \geq 0.180$, Fisher-Pitman permutation tests for paired replicates).

We then compared the pediatricians' decisions in both experimental groups. Table 2 shows differences in the length of antibiotic therapies between the second and first stages and between the third and second stages for pediatricians in both groups. In the intervention group, the pediatricians' mean change in the number of days of antibiotic treatment after announcement of feedback was –0.15 days (SD, 0.63; 95% CI, –0.34 to 0.06). The mean change in the number of days was –0.06 (SD, 0.25; 95% CI, –0.28 to 0.16) for pediatricians in the control group; this change did not differ significantly from the change in the intervention group ($P = 0.577$, Fisher-Pitman permutation test for independent samples). After feedback had been provided to pediatricians in the intervention group, the number of days of antibiotic treatment changed, on average, by –0.60 (SD, 0.97; 95% CI, –0.91 to –0.29). For pediatricians in the control group, the average change in the number of days in stage 3 was –0.06 (SD, 0.25; 95% CI, –0.15 to 0.03). The change in the intervention group was significantly larger than in the control group ($P = 0.000$, Fisher-Pitman permutation test for independent samples).

A study sample of 73, with 39 subjects in the intervention group and 34 subjects in the control group, gave the experiment a statistical power of 82% to detect an average effect of feedback in size of a reduction by 0.54 days (difference in changes from stage 2 to stage 3 between both groups) assuming a 2-tailed Mann-Whitney U test at a 5% significance level with an SD of 0.97 in the intervention group and an SD of 0.25 in the control group. A power analysis of the effect we defined as relevant a priori (a difference of 0.5 days between the groups) yielded an achieved power of 85% for a 2-tailed Mann-Whitney U test with a 5% significance level. For more details on the power analyses, see Online Appendix A.6.

To assess the effect of feedback on the appropriateness of therapy decisions, we analyzed the pediatricians' absolute deviation from the experts' recommended length of

Figure 2 The effect of feedback on the length of antibiotic therapies. This figure plots individual pediatricians' antibiotic therapy decisions (averaged over the 40 cases) for the 3 stages of the experiment in the intervention group. In each stage, 39 subjects decided on the length of antibiotic therapies for 40 routine medical cases, presented in random order on the subjects' computer screens. No feedback was given in the first stage; feedback was announced at the beginning of the second and third stages and shown after the second and third stages.

Table 2 Differences in Days of Antibiotic Therapy and Absolute Deviations from the Expert Recommendations

| | Experimental Group | | |
	Feedback (Intervention, $n = 39$)	No Feedback (Control, $n = 34$)	*P*-Value
A. Average changes in days of therapy			
$d_2 - d_1$	−0.15 (0.63)	−0.06 (0.63)	0.577
$d_3 - d_2$	−0.60 (0.97)	−0.06 (0.25)	0.000
B. Average changes in absolute deviation from the expert recommendations			
$\Delta_2 - \Delta_1$	−0.15 (0.56)	−0.09 (0.45)	0.587
$\Delta_3 - \Delta_2$	−0.33 (0.73)	0.00 (0.27)	0.004

Notes. This table shows average changes in days of antibiotic therapy and in absolute deviation from the expert recommendations for subjects in both experimental groups. Standard deviations are in parentheses. Note that d_t denotes days and Δ_t the average absolute deviation per subject from the expert recommendation B for cases $i = 1, 2, \ldots, 40$ and subjects $j = 1, 2, \ldots, J$ with $J \in \{34, 39\}$ in stage $t \in \{1, 2, 3\}$ of the experiment. More formally, $\Delta_t = \frac{1}{J} \frac{1}{40} \sum_{j=1}^{J} \sum_{i=1}^{40} |d_{ijt} - B_i|$ with $B_i = \frac{1}{20} \sum_{k=1}^{20} d_{ik}$ for experts $k = 1, 2, \ldots, 20$. *P*-values for differences between the groups are shown for 2-sided Fisher-Pitman permutation tests for independent samples. Wilcoxon-Mann-Whitney *U* tests yielded very similar *P*-values.

therapies; see Table 2. In stage 1, the pediatricians' absolute deviation from the experts' recommendations was not significantly different between the intervention and the control groups ($P = 0.301$, Fisher-Pitman permutation test for independent samples). In the intervention group, the difference between the pediatricians and the expert recommendations was weakly significantly affected by the announcement of feedback ($P = 0.085$,

Fisher-Pitman permutation test for paired replicates). After providing feedback, the deviation from the experts significantly decreased in the intervention group ($P=0.001$, Fisher-Pitman permutation test for paired replicates). In the control group, we observed no significant differences between the stages (both $P\geq0.278$, Fisher-Pitman test for paired replicates). Announcing feedback did not lead to more appropriate therapy decisions, as changes in deviation from stage 1 to stage 2 were not significantly different in the intervention and the control group ($P=0.587$, Fisher-Pitman permutation test for independent samples). From stage 2 to stage 3, the reduction in deviation was significantly greater in the intervention group than in the control group ($P=0.004$, Fisher-Pitman permutation test for independent samples). Wilcoxon signed-rank tests and Wilcoxon-Mann-Whitney U tests yielded very similar results.

Further, we used multilevel mixed-effects panel regressions to investigate the effect of feedback on antibiotic therapy decisions. For regression results, see Table 3. The effect of providing feedback is indicated by the interaction term "Third stage $\times$ Feedback". The effect of announcing feedback is indicated by the interaction term "Second stage $\times$ Feedback". The dependent variables are "days of antibiotic therapies" and "deviation from the expert recommendations", measured as the absolute difference between the pediatricians' decisions and the experts' recommended length of therapies.

The estimates of the regression models support the results of our nonparametric analyses. The provision of feedback in the intervention group led to a highly significant reduction in length of antibiotic therapies and absolute deviation from the recommendations, while the announcement had no statistically significant effect. These findings are robust when adding individual-specific controls, including sex, experience, personality traits, and economic preferences; for length of therapies, see Model 3; for absolute deviation from the expert recommendations, see Model 6 in Table 3. All models include session-, subject-, and case-specific random effects and account for potential within-group correlation. We conducted several analyses to check the robustness of our main results; see Online Appendix C.

Pediatricians' Characteristics and Antibiotic Prescribing

To analyze how pediatricians' characteristics relate to their antibiotic therapy decisions, we considered decisions made in the first stage of the experiment and merged data from the control and the intervention groups. The average length of antibiotic therapy for the 40 cases chosen in the first stage was of 7.53 days (95% CI, 7.32 to 7.73; $n=2,920$).

Table 4 shows estimation results from multilevel mixed-effects regression models. The pediatricians' experience was highly significantly associated with the length of antibiotic therapies. The longer pediatricians had practiced in a hospital, the shorter was the length of therapies and the smaller was the absolute deviation from the expert recommendations. Furthermore, the length of therapies significantly declined with the pediatricians' increasing willingness to take risks,[52–54] while the deviation from the expert recommendations was not significantly related to pediatricians' risk attitudes. Concerning personality traits,[50,51] we found that more conscientious pediatricians chose shorter therapies and, by doing so, deviated less from the experts. Other personality traits were not significantly associated with the pediatricians' decisions. In the regressions, we controlled for pediatricians' economic preferences, which comprised validated measures for trust, reciprocity, and altruism, and time preferences.[52–54]

Association Between Feedback and Pediatricians' Experience

Results from our regressions showed a consistent association between pediatricians' antibiotic prescribing decisions and experience: more experienced physicians chose shorter therapies and deviated less from the experts' recommendations. Using multilevel mixed-effects panel regression models, we tested whether the effect of feedback was specific to pediatricians' experience; see Table 5 for regression results. The positive coefficient of the interaction between the effect of feedback and experience suggests that pediatricians with less experience responded more strongly to feedback. More specifically, the less experienced the pediatricians were, the larger the effect of feedback was on the length of therapies and the appropriateness of antibiotic therapies. The effect of feedback decreased in the pediatricians' experience, suggesting that feedback does *not* mitigate the positive impact of experience on antibiotic prescribing decisions.

Discussion

We introduced a framed field experiment with pediatricians to analyze the causal effect of expert feedback on antibiotic prescribing in a tertiary pediatric care setting. Pediatricians decided on the length of antibiotic treatment for a series of routine pediatric cases. We found that

Table 3 Multilevel Mixed-Effects Panel Regression Models on the Effect of Feedback on Antibiotic Therapy Decisions

	Length of Antibiotic Therapies (in Days)			Absolute Deviation from the Expert Recommendations (in Days)		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Fixed effects						
Feedback (= 1 if intervention)		0.937** (0.423)	0.875 (0.560)		0.312 (0.294)	0.250 (0.364)
Second stage (= 1 if second stage)	-0.107 (0.073)	-0.063 (0.107)	-0.063 (0.107)	-0.122** (0.060)	-0.086 (0.088)	-0.086 (0.088)
Third stage (= 1 if third stage)	-0.450*** (0.108)	-0.112 (0.150)	-0.112 (0.150)	-0.297*** (0.082)	-0.085 (0.115)	-0.085 (0.115)
Effect of announcement (Second stage × Feedback)		-0.082 (0.147)	-0.082 (0.147)		-0.068 (0.120)	-0.068 (0.120)
Effect of feedback (Third stage × Feedback)		-0.633*** (0.205)	-0.633*** (0.205)		-0.397** (0.158)	-0.397** (0.158)
Female (= 1 if female)			0.581 (0.439)			0.073 (0.218)
Experience (years in hospital)			-0.077** (0.032)			-0.050*** (0.016)
Willingness to take risks			-0.213*** (0.081)			0.001 (0.040)
Extraversion			0.062 (0.141)			0.074 (0.070)
Agreeableness			-0.084 (0.203)			-0.051 (0.101)
Conscientiousness			-0.309 (0.209)			-0.278*** (0.104)
Neuroticism			-0.093 (0.134)			-0.025 (0.066)
Openness			0.150 (0.139)			-0.002 (0.069)
Further individual characteristics (economic preferences)	No	No	Yes	No	No	Yes
Constant	7.527*** (0.211)	7.035*** (0.311)	7.660*** (1.648)	2.919*** (0.152)	2.752*** (0.215)	4.543*** (0.839)
Random effects						
Session level						
Var(constant)	0.000 (0.000)	0.017 (0.112)	0.289 (0.283)	0.010 (0.043)	0.000	0.101*** (0.089)
Subject level						
Var(Stage 2)	0.222*** (0.065)	0.226*** (0.066)	0.226*** (0.066)	0.150*** (0.043)	0.153*** (0.044)	0.153*** (0.044)
Var(Stage 3)	0.686* (0.143)	0.595*** (0.128)	0.595* (0.128)	0.375*** (0.081)	0.342*** (0.076)	0.342*** (0.076)
Var(Constant)	2.567*** (0.556)	2.364*** (0.521)	1.583 (0.494)	1.405* (0.265)	1.396* (0.263)	1.094 (0.242)
Cov(Stage 2, Stage 3)	0.241*** (0.078)	0.232*** (0.075)	0.232*** (0.075)	0.147*** (0.048)	0.143*** (0.047)	0.143*** (0.047)
Cov(Stage 2, Constant)	-0.299** (0.141)	-0.285** (0.137)	-0.245* (0.137)	-0.277*** (0.085)	-0.276*** (0.086)	-0.286*** (0.084)
Cov(Stage 3, Constant)	-1.084*** (0.242)	-0.944*** (0.219)	-0.740*** (0.215)	-0.679*** (0.136)	-0.653*** (0.131)	-0.586*** (0.124)
Case level						
Var(Constant)	24.108*** (0.785)	24.108*** (0.669)	24.108*** (0.669)	4.732*** (0.145)	4.732*** (0.145)	4.732*** (0.145)
Var(Residual)	3.322*** (0.062)	3.322*** (0.062)	3.322*** (0.062)	2.192*** (0.041)	2.192*** (0.041)	2.192*** (0.041)
Number of observations	8760	8760	8760	8760	8760	8760
Number of subjects	73	73	73	73	73	73
Number of sessions	8	8	8	8	8	8

Notes. This table shows parameter estimates from multilevel mixed-effects REML regressions. The interaction "Third stage × Feedback" indicates the effect of showing feedback to subjects. In Models 1 to 3, the dependent variable is "length of antibiotic therapies (in days)." In Models 4 to 6, the dependent variable is "absolute deviation from the expert recommendations," measured in absolute values of the difference between the pediatricians' choices and the experts' recommended therapy length (in days). For each case, the subjects' choices were compared to the experts' aggregate opinion for the respective case. Standard errors are shown in parentheses. "Economic preferences" comprise validated measures for trust, reciprocity, and altruism, as well as time and risk preferences.[52–54] All models include session-, subject-, and case-specific random effects. In Model 5, the variance component at the session level is close to 0. Therefore, this model was estimated without grouping on the session level.
*P < 0.1, **P < 0.05, ***P < 0.01.

Table 4 Regressions on the Association of Antibiotic Therapy Decisions with Pediatricians' Characteristics

	Length of Antibiotic Therapies (in Days)		Absolute Deviation from the Expert Recommendations (in days)	
	Model 1		Model 2	
Fixed effects				
Female (= 1 if female)	0.856	(0.528)	0.394	(0.399)
Experience (years in hospital)	−0.110***	(0.039)	−0.076***	(0.030)
Willingness to take risks	−0.291***	(0.098)	−0.102	(0.074)
Extraversion	0.082	(0.169)	0.152	(0.128)
Agreeableness	0.154	(0.246)	−0.035	(0.185)
Conscientiousness	−0.581**	(0.252)	−0.538***	(0.190)
Neuroticism	0.159	(0.161)	0.136	(0.122)
Openness	0.205	(0.167)	0.165	(0.126)
Constant	8.912***	(2.001)	4.358***	(1.497)
Random effects				
Session level				
Var(Constant)	0.846	(0.583)	0.265	(0.217)
Subject level				
Var(Constant)	1.255	(0.408)	0.952	(0.235)
Case level				
Var(Constant)	25.651	(77.382)	6.750	(54.123)
Var(Residual)	3.342	(77.380)	0.998	(54.123)
Number of observations	2920		2920	
Number of subjects	73		73	
Number of sessions	8		8	

Notes. This table shows parameter estimates from multilevel mixed-effects REML regressions, considering the first stage of the experiment. The dependent variables are "length of antibiotic therapies" and "absolute deviation from the expert recommendations", both measured in days. Standard errors are shown in parentheses. "Willingness to take risks" was measured on a Likert scale ranging from 0 (fully risk averse) to 10 (fully risk seeking).[52–54] Besides the Big Five personality traits,[50,51] which are displayed in the table, we controlled for "economic preferences", which comprise validated measures for trust, reciprocity, and altruism, as well as risk and time preferences,[52–54] in both models. Both models include session-, subject-, and case-specific random effects.
*$P < 0.1$, **$P < 0.05$, ***$P < 0.01$.

providing pediatricians with simple directional expert feedback significantly reduced the length of antibiotic therapies by, on average, 10%. The absolute deviation of the pediatricians' decisions from length of therapies recommended by experts decreased significantly. The experimental data thus suggest that the expert benchmark "nudged" pediatricians toward a more *appropriate* use of antibiotics.[69]

The combination of experimental and survey data allowed us to relate pediatricians' decisions on the length of antibiotic therapies and their responses to feedback to their individual characteristics. We found that pediatricians who were more experienced and more conscientious chose shorter therapies and deviated significantly less from appropriate therapy durations. Previous studies in primary care settings, suggesting that more experienced physicians prescribe antibiotics more often,[47–49] neither considered the length, nor did they assess the

appropriateness of antibiotic therapies. While subjects in our experiment responded to feedback in a heterogeneous way, the main effect of feedback was robust toward the pediatricians' characteristics. When considering the interaction between the effect of feedback and experience, we found that feedback was most effective for physicians with little experience. These findings suggest that descriptive expert feedback can nudge pediatricians toward more appropriate antibiotic prescribing and compensates to some extent for a lack of experience.

The expert benchmark shown to pediatricians transmitted a descriptive normative message,[29] which was directed at pediatricians' self-image concerns.[70–72] Capitalizing on the human capacity to reflexive thinking,[73] this implies that expert feedback might have triggered pediatricians' self-directed concerns and may refer

Table 5 Multilevel Mixed-Effects Panel Regression Models on the Association between Individual Characteristics and Responses to Feedback

Characteristic	Length of Antibiotic Therapies (in Days)		Absolute Deviation from the Expert Recommendations (in Days)	
	Model 1		Model 2	
Fixed effects				
Feedback (= 1 if intervention)	0.801	(0.567)	0.205	(0.376)
Second stage (= 1 if second stage)	–0.063	(0.107)	–0.086	(0.088)
Third stage (= 1 if third stage)	–0.112	(0.148)	–0.085	(0.114)
Effect of announcement (Second stage × Feedback)	–0.082	(0.147)	–0.068	(0.120)
Effect of feedback (Third stage × Feedback)	–0.879***	(0.222)	–0.548***	(0.168)
Experience (years in hospital)	–0.109***	(0.034)	–0.074***	(0.019)
Experience × Effect of feedback	0.049***	(0.018)	0.030**	(0.013)
Female (= 1 if female)	0.499	(0.422)	0.022	(0.213)
Willingness to take risks	–0.208***	(0.078)	0.003	(0.039)
Constant	7.558***	(1.590)	4.566***	(0.825)
Random effects				
Session level				
Var(Constant)	0.307	(0.278)	0.120***	(0.098)
Subject level				
Var(Stage 2)	0.226***	(0.066)	0.153***	(0.044)
Var(Stage 3)	0.583**	(0.126)	0.332***	(0.074)
Var(Constant)	1.577	(0.482)	1.068	(0.234)
Cov(Stage 2, Stage 3)	0.233***	(0.075)	0.144***	(0.047)
Cov(Stage 2, Constant)	–0.239*	(0.136)	–0.283***	(0.083)
Cov(Stage 3, Constant)	–0.778***	(0.214)	–0.577***	(0.121)
Case level				
Var(Constant)	24.108***	(0.669)	4.732***	(0.145)
Var(Residual)	3.322***	(0.062)	2.192***	(0.041)
Number of observations	8760		8760	
Number of subjects	73		73	
Number of sessions	8		8	

Notes. This table shows parameter estimates from multilevel mixed-effects REML regressions. The interaction "Third stage × Feedback" indicates the effect of showing feedback to subjects. The interaction "Experience × Effect of feedback" indicates the association between the subjects' experience (number of years worked in hospital) and the effect of feedback. Standard errors are shown in parentheses. In both models, we controlled for the Big Five personality traits[50,51] and for "economic preferences", which comprise validated measures for trust, reciprocity, and altruism, as well as time and risk preferences.[52–54] Both models include session-, subject-, and case-specific random effects.
*$P < 0.1$, **$P < 0.05$, ***$P < 0.01$.

to the awareness of congruence between the expert benchmark and their own antibiotic therapy decisions. In contrast to self-image concerns, social-image concerns appear when people are observed by others and when their behavior is judged against a standard or a norm.[70–72] As the pediatricians took their decisions in anonymity and no information on identity or treatment patterns was shared among participants in the experiment, it seems more likely that changes in decisions were due to self-image concerns than to social-image concerns. Related experimental studies investigating the effect of social norm feedback and peer comparison rather focus on physicians' social-image concerns.[20,21,74] We add to this recent stream of the literature by providing evidence on the effect of descriptive expert feedback, which addresses physicians' self-image concerns in the absence of peer comparisons at an *individual* decision maker level. Furthermore, while prior research has mainly focused on whether and which antibiotics are prescribed,[16,20,21] we investigate the effect of a simple feedback mechanism on physicians' decisions regarding the *length* of antibiotic therapies.

Limitations and Future Research

We now discuss potential limitations of our study and avenues for future research. First, one might argue that the experimental design is somewhat simplistic to be reflective of a real clinical setting. It is true that we elicited hypothetical treatment decisions and that our experimental frame is parsimonious, used a set of hypothetical cases, and did not allow physicians to acquire additional information to assess the cases further. More specifically, one might argue that we were unable to consider pediatricians' responses to influencing factors, which are relevant in real-world clinical settings but might go beyond the constructed case descriptions (e.g., parental expectations[47] or risk and efficacy perceptions[75]). Second, different interpretations of the same case information may have affected individual physicians' judgments of the cases and their treatment decisions.[76,77] While we kept the information constant for all subjects, our study design and analyses did not focus on the process by which the therapy decisions were made (e.g., what heuristics had been used). Third, in real clinical practice, physicians may ask colleagues, search for information in guidelines, or order additional lab tests before making a treatment decision. The purpose of our experiment was to isolate the effect of feedback for a given, comparable set of information for each case. Not giving them the option to acquire additional information ensured that all pediatricians based their decisions on exactly the same information, allowing us to draw causal inferences on the effect of feedback. We made a real *ceteris paribus* variation of feedback, controlled the decision environment, and avoided confounding factors that potentially affect pediatricians' decisions.

Another concern might relate to the aggregated nature of our feedback mechanism. We employed an aggregated benchmark instead of case-by-case recommendations. By doing so, our experimental design allowed us to examine whether a "simple" feedback intervention raised awareness for appropriate use of antibiotics while maintaining the discretion for the pediatrician to decide on antibiotic prescribing for each medical case. After provision of feedback, we observed an overall change in the length of therapies toward what is more appropriate. One might argue that our aggregated results could conceal negative changes in therapy length for individual cases. Analyses on a case level, however, showed rather the opposite: for the vast majority of the cases, both the length of therapies and the absolute deviation from the expert recommendations decreased. Changes in the opposite direction for the remaining cases were not statistically significant; see Online Appendix C for details. Our results hence

suggest that comparison of own their prescribing decisions with an expert benchmark raises pediatricians' awareness for judicious use of antibiotics but maintains their individual and case-specific discretion when deciding whether and for which cases to adjust their treatment decisions.

An appealing feature of our parsimonious design is that it lends itself to further research. For example, future research could consider whether our findings can be translated to real clinical practice and to medical areas beyond pediatrics. Our findings also call for further studies investigating how long-lasting an effect of descriptive expert feedback on physicians' antibiotic prescribing decisions is and how the effect can be maintained. Another interesting question would be whether and, if so, how antibiotic prescribing decisions are affected by decision support tools that provide physicians with case-specific therapy recommendations and consider case-specific ranges of appropriate therapy durations. Relatedly, future studies could investigate the differential effect of expert and guideline-based feedback. Another potential avenue for future research would be to address social-image rather than self-image concerns.

Conclusion

Our experimental results suggest that descriptive expert feedback affects individual pediatricians' antibiotic prescribing decisions. Using a novel methodology and taking interindividual differences into account, we have shown that expert feedback, which conveys a normative message on antibiotic prescribing, can be an effective means in guiding pediatricians toward a more appropriate use of antibiotics. Most important, our results suggest that it is especially useful if targeted at physicians with low levels of experience. Our findings are also of practical importance as they provide an argument for the inclusion of individual feedback addressing the physicians' self-image in antibiotic stewardship programs.

ORCID iDs

Kyriakos Martakis (iD) https://orcid.org/0000-0003-3982-0914
Daniel Wiesen (iD) https://orcid.org/0000-0003-4627-1730

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://journals.sagepub.com/home/mdm.

References

1. Laxminarayan R, Duse A, Wattal C, et al. Antibiotic resistance—the need for global solutions. *Lancet Infect Dis*. 2013;13(12):1057–98.
2. Laxminarayan R, Amábile-Cuevas CF, Cars O, et al. UN high-level meeting on antimicrobials—what do we need? *Lancet*. 2016;388(10041):218–20.
3. Levy SB, Marshall B. Antibacterial resistance worldwide: causes, challenges and responses. *Nat Med*. 2004;10:122–9.
4. Nathan C, Cars O. Antibiotic resistance—problems, progress, and prospects. *N Engl J Med*. 2014;371:1761–3.
5. Haider BA, Lassi ZS, Bhutta ZA. Short-course versus long-course antibiotic therapy for non-severe community-acquired pneumonia in children aged 2 months to 59 months. *Cochrane Database Syst Rev*. 2008;2:CD005976.
6. Lassi ZS, Imdad A, Bhutta ZA. Short-course versus long-course intravenous therapy with the same antibiotic for severe community-acquired pneumonia in children aged two months to 59 months. *Cochrane Database Syst Rev*. 2017;10:CD008032.
7. Spellberg B. The new antibiotic mantra—"shorter is better." *JAMA Intern Med*. 2016;176(9):1254–5.
8. Hersh AL, Shapiro DJ, Pavia AT, Shah SS. Antibiotic prescribing in ambulatory pediatrics in the United States. *Pediatrics*. 2011;128(6):1053–61.
9. Downes KJ, Hahn A, Wiles J, Courter JD, Vinks AA. Dose optimisation of antibiotics in children: application of pharmacokinetics/pharmacodynamics in paediatrics. *Int J Antimicrob Agents*. 2014;43(3):223–30.
10. Schulman J, Dimand RJ, Lee HC, Duenas GV, Bennett MV, Gould JB. Neonatal intensive care unit antibiotic use. *Pediatrics*. 2015;135(5):826–33.
11. Hyun DY, Hersh AL, Namtu K, et al. Antimicrobial stewardship in pediatrics: how every pediatrician can be a steward. *JAMA Pediatr*. 2013;167(9):859–66.
12. Vodicka TA, Thompson M, Lucas P, et al. Reducing antibiotic prescribing for children with respiratory tract infections in primary care: a systematic review. *Br J Gen Pract*. 2013;63(612):e445–54.
13. Davey P, Marwick CA, Scott CL, et al. Interventions to improve antibiotic prescribing practices for hospital inpatients. *Cochrane Database Syst Rev*. 2017;2:CD003543.
14. Welschen I, Kuyvenhoven MM, Hoes AW, Verheij TJ. Effectiveness of a multiple intervention to reduce antibiotic prescribing for respiratory tract symptoms in primary care: randomised controlled trial. *BMJ*. 2004;329(7463):431–6.
15. Gjelstad S, Høye S, Straand J, Brekke M, Dalen I, Lindbæk M. Improving antibiotic prescribing in acute respiratory tract infections: cluster randomised trial from Norwegian general practice (prescription peer academic detailing (Rx-PAD) study). *BMJ*. 2013;347:f4403.
16. Gerber JS, Prasad PA, Fiks AG, et al. Effect of an outpatient antimicrobial stewardship intervention on broad-spectrum antibiotic prescribing by primary care pediatricians: a randomized trial. *JAMA*. 2013;309(22):2345–52.
17. Falk A, Heckman JJ. Lab experiments are a major source of knowledge in the social sciences. *Science*. 2009;326(5952):535–8.
18. Herbst D, Mas A. Peer effects on worker output in the laboratory generalize to the field. *Science*. 2015;350(6260):545–9.
19. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*. 2012;6:CD000259.
20. Hallsworth M, Chadborn T, Sallis A, et al. Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *Lancet*. 2016;387(10029):1743–52.
21. Meeker D, Linder JA, Fox CR, et al. Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: a randomized clinical trial. *JAMA*. 2016;315(6):562–70.
22. Harrison GW, List JA. Field experiments. *J Econ Lit*. 2004;42(4):1009–55.
23. Galizzi MM, Wiesen D. Behavioural experiments in health: an introduction. *Health Econ*. 2017;26(3):3–5.

24. Galizzi MM, Wiesen D. Behavioral experiments in health economics. In: Hamilton JH, Dixit A, Edwards S, Kenneth J, eds. *Oxford Research Encyclopedia of Economics and Finance*. Oxford, UK: Oxford University Press; 2018.

25. Clee M, Wicklund R. Consumer behavior and psychological reactance. *J Consum Res*. 1980;6(4):389–405.

26. Alcott H. Social norms and energy conservation. *J Public Econ*. 2011;95(9–10):1082–95.

27. Linder JA. Moving the mean with feedback: insights from behavioral science. *NPJ Prim Care Respir Med*. 2016;26: 16018.

28. Cialdini R, Reno R, Kallgren C. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J Pers Soc Psychol*. 1990;58(6): 1015–26.

29. Schultz PW, Nolan JM, Ciladini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. *Psychol Sci*. 2007;18(5):429–34.

30. Holtgrave DR, Lawler F, Spann SJ. Physicians' risk attitudes, laboratory usage, and referral decisions: the case of an academic family practice center. *Med Decis Making*. 1991;11(2):125–30.

31. Pearson SD, Goldman L, Orav EJ, et al. Triage decisions for emergency department patients with chest pain. *J Gen Intern Med*. 1995;10(10):557–64.

32. Allison JJ, Kiefe CI, Cook EF, Gerrity MS, Orav EJ, Centor R. The association of physician attitudes about uncertainty and risk taking with resource use in a Medicare HMO. *Med Decis Making*. 1998;18(3):320–9.

33. Michel-Lepage A, Ventelou B, Nebout A, Verger P, Pulcini C. Cross-sectional survey: risk-averse French GPs use more rapid-antigen diagnostic tests in tonsillitis in children. *BMJ Open*. 2013;3(10):e003540.

34. Nebout A, Cavillon M, Ventelou B. Comparing GPs' risk attitudes for their own health and for their patients': a troubling discrepancy? *BMC Health Serv Res*. 2018;18:283.

35. Massin S, Ventelou B, Nebout A, Verger P, Pulcini C. Cross-sectional survey: risk-averse French general practitioners are more favorable towards influenza vaccination. *Vaccine*. 2015;33(5):610–4.

36. Franks P, Williams GC, Zwanziger J, Mooney C, Sorbero M. Why do physicians vary so widely in their referral rates? *J Gen Intern Med*. 2000;15(3):163–8.

37. Forrest CB, Nutting PA, Von Schrader S, Rohde C, Starfield B. Primary care physician specialty referral decision making: patient, physician, and health care system determinants. *Med Decis Making*. 2006;26(1):76–85.

38. Rochon PA, Gruneir A, Bell CM, et al. Comparison of prescribing practices for older adults treated by female versus male physicians: a retrospective cohort study. *PLoS One*. 2011;13(10):e0205524.

39. Wallis CJ, Ravi B, Coburn N, Nam RK, Detsky AS, Satkunasivam R. Comparison of postoperative outcomes among patients treated by male and female surgeons: a population based matched cohort study. *BMJ*. 2017;359:j4366.

40. Tsugawa Y, Jena AB, Figueroa JF, Orav EJ, Blumenthal DM, Jha AK. Comparison of hospital mortality and readmission rates for Medicare patients treated by male vs female physicians. *JAMA Intern Med*. 2017;177(2):206–13.

41. Borghans L, Druckworth AL, Heckman JJ, ter Weel B. The economics and psychology of personality traits. *J Hum Resour*. 2008;43(4):972–1059.

42. Almlund M, Duckworth AL, Heckman J, Kautz T. Personality psychology and economics. In: Hanushek EA, Machin S, Woessmann L, eds. *Handbook of the Economics of Education*. Vol. 4. Amsterdam: North Holland; 2011. p 4–28.

43. Callen M, Gulzur S, Hasanain A, Khan Y, Rezaee A. *Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan*. NBER Working Paper Series No. 21180. National Bureau of Economic Research; 2015. Cambridge, MA.

44. Donato K, Miller G, Mohanan M, Truskinovsky Y, Vera-Hernández M. Personality traits and performance contracts: evidence from a field experiment among maternity care providers in India. *Am Econ Rev*. 2017;107(5):506–10.

45. Hulscher ME, Grol RP, van der Meer JW. Antibiotic prescribing in hospitals: a social and behavioural scientific approach. *Lancet Infect Dis*. 2010;10(3):167–75.

46. Rodrigues AT, Roque F, Falcão A, Figueiras A, Herdeiro MT. Understanding physician antibiotic prescribing behaviour: a systematic review of qualitative studies. *Int J Antimicrob Agents*. 2013;41(3):203–12.

47. Sirota M, Round T, Samaranayaka S, Kostopoulou O. Expectations for antibiotics increase their prescribing: causal evidence about localized impact. *Health Psychol*. 2017;36(4):402–9.

48. de Jong BM, van der Ent CK, van der Zalm MM, van Putte-Katier N, Verheij TJ, Kimpen JL, Uiterwaal CS. Respiratory symptoms in young infancy: child, parent and physician related determinants of drug prescription in primary care. *Pharmacoepidemiol Drug Saf*. 2009;18(7): 610–8.

49. Akkerman AE, Kuyvenhoven MM, Van der Wouden JC, Verheij TJ. Prescribing antibiotics for respiratory tract infections by GPs: management and prescriber characteristics. *Br J Gen Pract*. 2005;55(511):114–8.

50. Gosling SD, Rentfrow PJ, Swann WB Jr. A very brief measure of the Big Five personality domains. *J Res Pers*. 2003;37(6):504–28.

51. Rammstedt B, John OP. Measuring personality in one minute or less: a 10-item short version of the Big Five inventory in English and German. *J Res Pers*. 2007;41(1):203–12.

52. Falk A, Becker A, Dohmen T, Enke B, Huffman DB, Sunde U. Global evidence on economic preferences. *Q J Econ*. 2018;133(4):1645–92.

53. Falk A, Becker A, Dohmen T, Huffman DB, Sunde U. *The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences*. IZA Discussion Papers No. 9674. Institute for the Study of Labor; 2016. Bonn, Germany.

54. Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG. Individual risk attitudes: measurement, determinants, and behavioral consequences. *J Eur Econ Assoc.* 2011;9(3):522–50.

55. Chomsky N.A Review of B.F. Skinner's verbal behavior. *Language.* 1959;35:26–58.

56. Popper KR. *Objective Knowledge: An Evolutionary Approach.* Oxford, UK: Oxford University Press; 1972.

57. Felin T, Foss NJ. The endogenous origins of experience, routines, and organizational capabilities: the poverty of stimulus. *J Institut Econ.* 2011;7(2):231–56.

58. Levinthal DA, March J. The myopia of learning. *SMJ.* 1993;14:95–112.

59. Nonaka L. A dynamic theory of organizational knowledge creation. *Organ Sci.* 1994;5(1):14–37.

60. Olofsson SK, Cars O. Optimizing drug exposure to minimize selection of antibiotic resistance. *Clin Infect Dis.* 2007;45(Suppl 2):129–36.

61. Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. *JAMA.* 2001;285(22):2871–9.

62. Deutsche Gesellschaft für Pädiatrische Infektiologie (DGPI). *DGPI-Handbuch Infektionen bei Kindern und Jugendlichen.* 7th ed. Stuttgart, Germany: Thieme; 2018.

63. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PAC, Rubin HR. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA.* 1999;282(15):1458–65.

64. Barlow G, Nathwani D, Myers E, et al. Identifying barriers to the rapid administration of appropriate antibiotics in community acquired pneumonia. *J Antimicrob Chemother.* 2008(2);61:442–51.

65. May L, Gudger G, Armstrong P, et al. Multisite exploration of clinical decision making for antibiotic use by emergency medicine providers using quantitative and qualitative methods. *Infect Control Hosp Epidemiol.* 2014;35(9):1114–25.

66. Fischbacher U. Z-tree: Zurich toolbox for readymade economic experiments—experimenter's manual. *Exp Econ.* 2007;10(2):171–8.

67. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007;39(2):175–91.

68. List JA. Why economists should conduct field experiments and 14 tips for pulling one off. *J Econ Perspect.* 2011;25(3):3–16.

69. Thaler R, Sunstein CR. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* New Haven, CT: Yale University Press; 2008.

70. Bénabou R, Tirole J. Incentives and prosocial behavior. *Am Econ Rev.* 2006;96(5):1652–78.

71. Bénabou R, Tirole J. Identity, morals, and taboos: beliefs as assets. *Q J Econ.* 2011;126(2):805–55.

72. Falk A. *Facing Yourself—A Note on Self-Image.* CESifo Working Paper Series No. 6428. Center for Economic Studies and Ifo Institute (CESifo); Munich, Germany.

73. Leary MR, Tangney JP. The self as an organizing construct in the behavioral and social sciences. In: Leary MR, Tangney JP, eds. *Handbook of Self and Identity.* 2nd ed. New York: Guilford; 2012. p 1–18.

74. Meeker D, Knight TK, Friedberg MW, et al. Nudging guideline-concordant antibiotic prescribing: a randomized clinical trial. *JAMA Intern Med.* 2014;174(3):425–31.

75. Broniatowski DA, Klein EY, May L, Martinez EM, Ware C, Reyna VF. Patients' and clinicians' perceptions of antibiotic prescribing for upper respiratory infections in the acute care setting. *Med Decis Making.* 2018;38(5):547–61.

76. Kostopoulou O, Russo JE, Keenan G, Delaney BC, Douiri A. Information distortion in physicians' diagnostic judgments. *Med Decis Making.* 2012;32(6):831–9.

77. Nurek M, Kostopoulou O, Hagmayer Y. Predecisional information distortion in physicians' diagnostic judgments: strengthening a leading hypothesis or weakening its competitor? *Judgm Decis Mak.* 2014;9(6):572–85.