PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Mystery shopping and coaching as a form of audit and feedback to improve community pharmacy management of non-prescription medicine requests: an intervention study
AUTHORS	Collins, Jack; Schneider, Carl; Naughtin, Clare; Wilson, Frances; de Almeida Neto, Abilio; Moles, Rebekah

VERSION 1 – REVIEW

REVIEWER	Prof. Dr. Bernhard Langer Faculty of Health, Nursing, Administration University of Applied Sciences Neubrandenburg 17033 Neubrandenburg GERMANY
REVIEW RETURNED	24-Sep-2017

| GENERAL COMMENTS | This is an excellent intervention study exploring the impact of feedback to improve community pharmacy management in Australia. I think this will be a strong addition to the literature. I nonetheless wish to make a few comments and recommendations.

1. P. 5 line 39: 6 instead of six.

2. P. 5 line 39: 2 instead of two.

3. P.5 line 37-39: Suggest a very brief mention why 6 pharmacies did not repeat their participation.

4. P. 5 line 51-58: Suggest to make it clearer that the study is not covert.

5. P. 6 line 24-30: Suggest to better describe the allocation of the students to the pharmacies. How many times does a student visit a specific pharmacy? This aspect is important because of the risk being identified visiting a specific pharmacy several times.

6. P. 6 line 24-30: Suggest to better describe the allocation of the pharmacies to the scenarios. Please describe in how many pharmacies each scenario was presented. Perhaps this information can be given additionally in table 1 (additionally to the number of visits). It is important because not all scenarios are presented in all 36 pharmacies.

7. P. 7 line 26: Please explain better how to reach the maximum questioning score of 18 points (perhaps it is better to describe here also the maximum score for the scenario ten (page 22). |

8. P. 7 line 49: Suggest to better describe here scenario ten (not the adult cough/cold scenario), because scenario ten is in the appendix (p. 21/22).

9. P. 8 line 48: It is reported that an independent variable "if written information was provided" is included in the regression model", but in table 2 this variable is missing. Why?

10. P. 9 line 6: I think it is necessary to mention additionally how many pharmacies were visited per scenario on average (on average 6 pharmacies (?) x 9 visits x 10 scenarios = 540 visits overall). As a result the 540 visits overall are better comprehensible.

11. P. 9 line 22: In 30 visits students were identified as mystery shoppers. In P. 9 line 12 you describe that in 8 visits the students were identified as mystery shoppers resulting in termination of the interaction. My conclusion is that in 22 visits students were identified as mystery shoppers not resulting in termination of the interaction. Is it right? Why does this not resulting in termination of the interaction? Please describe it a little bit more detailed.

12. P. 9 line 42-43: It is better to report first the pediatric fever scenario and then the smoking cessation scenario (in a similar way to table 1).

13. P. 10 line 3: I suggest writing "appropriate outcome" instead of "appropriateness", because it fits better to the phrases in the text.

14. P. 13 line 40: Yes, that's true, but I suggest to make it clearer that these differences (direct-product vs. symptom-based requests), however, cannot explain the differences within the scenarios of this study, because all scenarios are designed as direct-product requests.

15. P. 13 line 45: You can add here reference 22, because Horvath et al. reported similar results.

16. P. 13 line 51: Yes, that's true, there are some examples for these findings in the literature. But there are also some contrary study results in the international literature (see for example):

Horvat N et al. Using the Simulated Patient Methodology to Assess Paracetamol-Related Counselling for Headache. PLOS ONE 2013; 7(12): e52510

Kashyap KC et al. Management of over-the-counter insomnia complaints in Australian community pharmacies: a standardized patient study. Int J Pharm Pract 2014; 22(2): 125-134. doi:10.1111/ijpp.12052.

Therefore I suggest reporting that there are contradictory study results in the international literature in this regard.

17. P. 15 line 26: It is true that one limitation is that the study design is not covert and therefore the staff were aware that the study was taking place. On P. 6 line 5 you write that the pharmacies were not informed exactly when the visits were occur.

	But because of the study design (nine visits in nine weeks) and immediate feedback after the first visit (and the information for the pharmacies before the beginning of the study that they would receive one visit each week), the pharmacies knew after the first feedback that the next visit(s) will take place in a few days. I think that this enhances the hawthorne effect enormously (in contrast to a study design without immediate feedback). On the other hand immediate feedback is a recommendable coaching technique (see Xu et al. (2012)), but perhaps it would have been better to give feedback only after the last visit. I suggest discussing these aspects additionally in the limitations. 18. In general: In the paper you use sometimes the wording "appropriate outcome", sometimes the wording "key outcome". For a better understanding I suggest a standardization.

REVIEWER	Bridget Paravattil Qatar University Qatar
REVIEW RETURNED	26-Sep-2017

GENERAL COMMENTS	Comments for the authors: Mystery shopping and coaching as a form of audit and feedback to improve community pharmacy management of non-prescription medicine requests: an intervention study Abstract □ Line 24 – 521 – should be spelled out. Also this is not a complete sentence. Methods □ Page 5, Line 29 – what was the role of the BPharm student in the study. It should state that they were recruited to be mystery shoppers. □ Page 5, Line 33 – Did the students recruit the community pharmacy? Was there a criteria for them to follow when selecting the pharmacies? □ Page 6, Line 10 – Did students have to learn all scenarios or were they assigned to one scenario? □ Page 6, Line 34 – Scenario variables were altered at each visit. Who altered the scenario (student?) and what was the mechanism that was used to alter the each scenario? □ Page 7, Line 5 – what was the extent of the feedback to the staff member by the student? Was it a simple you did this and you did not do this or was there a detailed description on why it was an incorrect/correct response? Were students capable of providing detailed explanations for incorrect responses especially since they are still in the learning process themselves? Would the staff member take the feedback seriously when it was coming from a student? □ Page 7, Line 3 – staff members were provided with self-evaluation form. What was done with this information? Was it used during the feedback session?

		☐ Page 7, Line 43 – it would be nice to have a table with a brief description of the appropriate outcome for each of the scenarios.

☐ Page 7, Line 43 – it would be nice to have a table with a brief description of the appropriate
outcome for each of the scenarios.
Results
☐ Page 9, Line 12 – it states 8 interactions the mystery shopper was identified and them in Line 22
it states that 30 cases the students were identified as shoppers. Why were the 30 cases not
excluded? What is the difference between the 8 interactions and the 30 interactions?
☐ Page 9, Line 38 and onwards – the p values should be listed in table 1.
☐ Page 9, Line 44 – these values are not seen table 2. Did the authors mean table 1? Also should
it be Table 1 (not Table 2) also outline the proportions of visits…….
☐ Page 10, Line 33-43 – perhaps this information could be presented in a table. As one of the aims
of the study was if scores and outcomes were achieved over time. There are no tables showing
the questioning scores for the visits. It would be nice to see where the improvement occurred
whether it be in the questioning score or the outcome or for each scenario.
Discussion
☐ Page 13, Line 7-29 – every sentence here is cited. This paragraph needs to be revised to relate to
the study rather than give a description of FIT. As mentioned earlier - Were students capable of
providing detailed explanations for incorrect responses especially since they are still in the
learning process themselves? Would the staff member take the feedback seriously when it was
coming from a student? Could this possibly affect some of the scores and/or outcomes?
☐ I feel there should have been a section in the discussion related to the percentage of
pharmacists that achieved the outcome. Was this an acceptable percentage? Are the authors
content with results of the questioning scores and outcomes in relation to current practice in
Australia?
☐ How can the authors say that feedback improved when there is no guarantee that a mystery
shopper visited the same staff on a different occasion? Each of these encounters could have
been with different staff member. How many staff encountered more than one mystery
shopper? Can the authors truly say the feedback help in the next encounter?
☐ As a suggestion for future research it would be important

Reviewer 1: Professor Bernhard Langer

(1) P. 5 line 39: 6 instead of six.

(2) P. 5 line 39: 2 instead of two.

Response: Based on convention that numbers less than ten are written in words, we have left these unchanged. Furthermore on page 5 line 39, Two is the beginning of a sentence, this has been left unchanged. If the editor wishes to change words to figures we of course are ok with this convention.

(3) P.5 line 37-39: Suggest a very brief mention why 6 pharmacies did not repeat their participation.

Response: In conjunction with comment (21) from reviewer 2, this section has been re-worded to better explain the recruitment process. The six pharmacies did not take part as they were not in the first 24 pharmacies to reply to participating again in semester two. Reasons for this were not given to the research team.

"Fifty-nine third-year students consented to take part in the study, with 30 taking part in semester one of the academic year (March-June) and the remaining 29 taking part in semester two (July-November). Two fourth-year honours students shared the role of the remaining place in the second semester to give a total of 61 students taking part as mystery shoppers. Thirty-six of 59 third-year students successfully recruited one community pharmacy (n=36) in the Sydney metropolitan region (no other criteria for participating pharmacies were applied other than geographic location). Thirty of these pharmacies were selected to take part in semester one, with the remaining six reserved for semester two. The first 24 pharmacies from semester one to reply to an invitation to take part again in semester two were re-recruited, giving a total of 36 pharmacies taking part across both semesters."

(4) P. 5 line 51-58: Suggest to make it clearer that the study is not covert.

Response: "Thus, the study was not truly covert." Has been added to this section to make it clearer that this was the case.

(5) P. 6 line 24-30: Suggest to better describe the allocation of the students to the pharmacies. How many times does a student visit a specific pharmacy? This aspect is important because of the risk being identified visiting a specific pharmacy several times.

Response: Each student visited a particular pharmacy once, therefore the following sentence has been restructured in order to make this point clearer. "Each week individual students visited a different pharmacy in order to minimise the risk of detection."

(6) P. 6 line 24-30: Suggest to better describe the allocation of the pharmacies to the scenarios. Please describe in how many pharmacies each scenario was presented. Perhaps this information can be given additionally in table 1 (additionally to the number of visits). It is important because not all scenarios are presented in all 36 pharmacies.

Response: This section has been restructured for further clarification, "Each week individual students visited a different pharmacy in order to minimise the risk of detection. Each pharmacy was allocated a different scenario (one of 10) relating to a minor ailment. In each semester, three pharmacies were allocated each scenario (10 scenarios × 6 pharmacies × 9 visits = 540 visits in total).

Pharmacies that participated in both semesters were allocated a different scenario in the second semester to the one that was allocated in the first, meaning each pharmacy was allocated a maximum of two of 10 scenarios across the course of the study."

(7) P. 7 line 26: Please explain better how to reach the maximum questioning score of 18 points (perhaps it is better to describe here also the maximum score for the scenario ten (page 22).

Response: "For example, in the version of the asthma scenario included in the supplementary file, if a staff member asked all seven questions correctly they would be allocated a questioning score of 14 points, the maximum possible for that scenario." This has been included to aid understanding of how the scoring system works.

(8) P. 7 line 49: Suggest to better describe here scenario ten (not the adult cough/cold scenario), because scenario ten is in the appendix (p. 21/22).

Response: This has been re-worded to use the asthma example.

"For example, in the asthma scenario (online supplementary file 1) the staff member should have identified the patient's frequent use of salbutamol indicating poorly controlled asthma and appropriately referred the patient to a medical practitioner."

(9) P. 8 line 48: It is reported that an independent variable "if written information was provided" is included in the regression model", but in table 2 this variable is missing. Why?

Response: Thank you for detecting this error on our behalf. An earlier version of the regression model included this variable but ultimately this was dropped from the final model and has now been removed from the manuscript.

(10) P. 9 line 6: I think it is necessary to mention additionally how many pharmacies were visited per scenario on average (on average 6 pharmacies (?) x 9 visits x 10 scenarios = 540 visits overall). As a result the 540 visits overall are better comprehensible.

Response: Thank you for this point, please see response to comment (6) where this suggestion has been added to the methods section.

(11) P. 9 line 22: In 30 visits students were identified as mystery shoppers. In P. 9 line 12 you describe that in 8 visits the students were identified as mystery shoppers resulting in termination of the interaction. My conclusion is that in 22 visits students were identified as mystery shoppers not resulting in termination of the interaction. Is it right? Why does this not resulting in termination of the interaction? Please describe it a little bit more detailed.

Response: Your interpretation of this is correct, it was not until the completion of those 22 visits that the staff member revealed to the student they suspected they were a mystery shopper during the interaction. This section has been re-worded for clarification. "Of these 30 cases, in 22 instances it was not until the completion of the visit during the staff feedback that the staff revealed to the students that they had suspected they were a mystery shopper." Please also see response to comment (27).

(12) P. 9 line 42-43: It is better to report first the pediatric fever scenario and then the smoking cessation scenario (in a similar way to table 1).

Response: Thank you, this order has been updated.

(13) P. 10 line 3: I suggest writing "appropriate outcome" instead of "appropriateness", because it fits better to the phrases in the text.

Response: Thank you, I have made this suggested change.

(14) P. 13 line 40: Yes, that's true, but I suggest to make it clearer that these differences (direct-product vs. symptom-based requests), however, cannot explain the differences within the scenarios of this study, because all scenarios are designed as direct-product requests.

Response: Thank you, you make an excellent point. This section has been updated. "…which may explain the poor results seen in some scenarios, but does not explain inter-scenario differences."

(15) P. 13 line 45: You can add here reference 22, because Horvath et al. reported similar results.

Response: Thank you, this reference has been added.

(16) P. 13 line 51: Yes, that's true, there are some examples for these findings in the literature. But there are also some contrary study results in the international literature (see for example):

Horvat N et al. Using the Simulated Patient Methodology to Assess Paracetamol-Related Counselling for Headache. PLOS ONE 2013; 7(12): e52510

Kashyap KC et al. Management of over-the-counter insomnia complaints in Australian community pharmacies: a standardized patient study. Int J Pharm Pract 2014; 22(2): 125-134. doi:10.1111/ijpp.12052.

Therefore I suggest reporting that there are contradictory study results in the international literature in this regard.

Response: Thank you, this has been restructured as suggested with appropriate references included, "Existing literature in this field reports contradictory results in regard to a pharmacist being involved in the interaction and a correlation with a higher questioning score and subsequently an appropriate outcome. [11, 22, 34]"

(17) P. 15 line 26: It is true that one limitation is that the study design is not covert and therefore the staff were aware that the study was taking place. On P. 6 line 5 you write that the pharmacies were not informed exactly when the visits were occur. But because of the study design (nine visits in nine weeks) and immediate feedback after the first visit (and the information for the pharmacies before the beginning of the study that they would receive one visit each week), the pharmacies knew after the first feedback that the next visit(s) will take place in a few days. I think that this enhances the Hawthorne effect enormously (in contrast to a study design without immediate feedback). On the other hand immediate feedback is a recommendable coaching technique (see Xu et al. (2012)), but perhaps it would have been better to give feedback only after the last visit. I suggest discussing these aspects additionally in the limitations.

Response: Thank you for this comment, this section has been modified to address the fact that, yes, the timeframe of the visits would have become clear to the pharmacy during the course of the intervention. "Likewise, the staff were aware that the study was taking place, and despite the staff not knowing the exact timing of the visits, the likelihood of future visits would have become apparent after the initial visit and feedback session."

This intervention was designed to utilise repetition of audit and feedback to elicit behaviour change; the Hawthorne effect is thereby used to enhance the utility of the intervention. The benefit of an induced Hawthorne effect has been added to the third paragraph of the discussion.

The discussion has been updated to include the notion of being "watched" or knowing that mystery shoppers may be coming to the pharmacy at any time facilitates a state of "hypervigilance" for the pharmacy staff whereby they perform at their best level of practice. Anticipation of a mystery shopping visit is a driver for improved practice over the course of the visit schedule and is a benefit of the methodology we have employed in this study.

(18) In general: In the paper you use sometimes the wording "appropriate outcome", sometimes the wording "key outcome". For a better understanding I suggest a standardization.

Response: Thank you for identifying this inconsistency, all wording has been changed to "appropriate outcome".

Reviewer 2: Bridget Paravattil

Abstract
(19) Line 24 – 521 – should be spelled out. Also this is not a complete sentence.

Response: This has been updated. "Five hundred and twenty-one visits were analysed, of which 54% resulted in an appropriate outcome"

Methods
(20) Page 5, Line 29 – what was the role of the BPharm student in the study. It should state that they were recruited to be mystery shoppers

Response: This has been updated, thank you. "Bachelor of Pharmacy (BPharm) students entering their third year of the degree program at the University of Sydney in 2015 were invited to take part in the study to be mystery shoppers in lieu of a portion of their regular clinical placements program"

(21) Page 5, Line 33 – Did the students recruit the community pharmacy? Was there a criteria for them to follow when selecting the pharmacies?

Response: This has been updated for clarification, thank you. "Fifty-nine third-year students consented to take part in the study, with 30 taking part in semester one of the academic year (March-June) and the remaining 29 taking part in semester two (July-November). Two fourth-year honours students shared the role of the remaining place in the second semester to give a total of 61 students taking part as mystery shoppers. Thirty-six of 59 third-year students successfully recruited one community pharmacy (n=36) in the Sydney metropolitan region (no other criteria for participating pharmacies were applied other than geographic location). Thirty of these pharmacies were selected to take part in semester one, with the remaining six reserved for semester two. The first 24 pharmacies from semester one to reply to an invitation to take part again in semester two were re-recruited, giving a total of 36 pharmacies taking part across both semesters."

(22) Page 6, Line 10 – Did students have to learn all scenarios or were they assigned to one scenario?

Response: This has been updated to clearly state students were trained in all scenarios. "Students were trained across a two-day program where they learned how to enact all 10 scenarios (of which they went on to shop nine) through role-play with the research team, familiarised themselves with data collection sheets, were taught about theory behind the intervention, and trained in providing feedback to pharmacy staff."

(23) Page 6, Line 34 – Scenario variables were altered at each visit. Who altered the scenario (student?) and what was the mechanism that was used to alter the each scenario?

Response: The scenarios were all written in advance of the study commencing by the researchers, this sentence has been altered to make it clearer that this was not at the student's discretion. "Scenarios variables were altered each visit by the research team in the script given to students, including…"

(24) Page 7, Line 5 – what was the extent of the feedback to the staff member by the student? Was it a simple you did this and you did not do this or was there a detailed description on why it was an incorrect/correct response? Were students capable of providing detailed explanations for incorrect responses especially since they are still in the learning process themselves? Would the staff member take the feedback seriously when it was coming from a student?

Response: Thank you for this comment and your questions. The students were able to provide feedback primarily through a facilitated self-reflection as used by Xu et al (2012). In addition, students would provide varying levels of other general feedback depending on the circumstances of the interaction, such as the willingness of the individual to receive feedback and the busyness of the pharmacy. Students were asked to research the scenarios before undertaking the visits and the content of the scenarios aligned with content they had either already studied or were currently studying. As most scenarios were designed as common over-the-counter requests it is not unreasonable to expect students to have an understanding of the issue in sufficient detail to explain this to a staff member. Anecdotally, staff were asked for feedback at the completion of the study and said that they, particularly pharmacy assistants, found the process valuable and did actually learn some contemporary practice points from students.

(25) Page 7, Line 3 – staff members were provided with self-evaluation form. What was done with this information? Was it used during the feedback session?

Response: The staff members filled out their evaluation form as a form of self-reflection so they could identify where they may have had some shortcomings during the interaction. The completed staff evaluation forms have been collected and will be analysed for the purpose of another manuscript. "Within five minutes students returned to the pharmacy, provided the staff member with a blank copy of the scoresheet for them to complete as a form of self-evaluation and then provided verbal feedback to the staff member based on their performance. Any products purchased were returned to the pharmacy without refund and self-evaluation forms were collected from the staff member."

(26) Page 7, Line 43 – it would be nice to have a table with a brief description of the appropriate outcome for each of the scenarios.

Response: Thank you for this comment, a table with these outcomes has been added as suggested.

Results
(27) Page 9, Line 12 – it states 8 interactions the mystery shopper was identified and them in Line 22 it states that 30 cases the students were identified as shoppers. Why were the 30 cases not excluded? What is the difference between the 8 interactions and the 30 interactions?

Response: Thank you for this comment. In accordance with the other reviewer's comments this section has been updated to make this clearer. In eight situations the staff member terminated the interaction before completion, whereas the remaining 22 proceeded through to feedback where the staff member stated that they suspected the student was a mystery shopper. This section has been re-worded to read: "Students were identified as mystery shoppers in 6% of cases (n=30). Of these 30 cases, in 22 instances it was not until the completion of the visit during the staff feedback that the staff revealed to the students that they had suspected they were a mystery shopper."

In this mystery shopping study we did not intend the visits to be a merely an measure of practice, but rather as an intervention in conjunction with feedback to improve practice. Repetition of the visits and feedback results in anticipation or a state of "hypervigilance". In order to measure this effect we must include visits where the shopper was suspected to be a mystery shopper, but was not disclosed until the completion of the visit. The regression model demonstrated that staff performance did not significantly differ in instances where students were identified as mystery shoppers, thereby suggesting that the identification of the mystery shopper does not alter performance and supporting the concept of "hypervigilance".

(28) Page 9, Line 38 and onwards – the p values should be listed in table 1.

Response: Thank you for this comment, the p values have been added to table 1 (now table 2).

(29) Page 9, Line 44 – these values are not seen table 2. Did the authors mean table 1? Also should it be Table 1 (not Table 2) also outline the proportions of visits…….

Response: Thank you for identifying this error, this has been corrected.

(30) Page 10, Line 33-43 – perhaps this information could be presented in a table. As one of the aims of the study was if scores and outcomes were achieved over time. There are no tables showing the questioning scores for the visits. It would be nice to see where the improvement occurred whether it be in the questioning score or the outcome or for each scenario.

Response: Thank you for this point, an additional table has been added with questioning scores similar to the table reporting appropriate outcome over time.

Discussion
(31) Page 13, Line 7-29 – every sentence here is cited. This paragraph needs to be revised to relate to the study rather than give a description of FIT. As mentioned earlier - Were students capable of providing detailed explanations for incorrect responses especially since they are still in the learning process themselves? Would the staff member take the feedback seriously when it was coming from a student? Could this possibly affect some of the scores and/or outcomes?

Response: Thank you for this comment, we feel that an explanation of FIT is necessary for this discussion point in order to improve reader understanding of the concept. Further comments have been added to this paragraph regarding the quality and acceptability of feedback provided by student mystery shoppers.

(32) I feel there should have been a section in the discussion related to the percentage of pharmacists that achieved the outcome. Was this an acceptable percentage? Are the authors content with results of the questioning scores and outcomes in relation to current practice in Australia?

Response: Thank you for this comment. Ideally according to practice standards, we would like to see every interaction with a pharmacy staff member to result in the appropriate counselling, advice, product selection, etc. There is no consensus or guideline on what determines an "acceptable percentage". The results in this study are not dissimilar to similar studies previously conducted in Australia. It was promising to see that over time pharmacies improved in both scores and outcomes, and that despite some initially poor results, these can be improved. We do not feel that adding our subjective opinion on this matter to the discussion adds value to the manuscript.

(33) How can the authors say that feedback improved when there is no guarantee that a mystery shopper visited the same staff on a different occasion? Each of these encounters could have been with different staff member. How many staff encountered more than one mystery shopper? Can the authors truly say the feedback help in the next encounter?

Response: Thank you for raising these valid points. The strengths and limitations section has been expanded to include these limitations of the study. We were not able to record the individual staff member who served the shopper each time due to ethics constrictions, meaning that we cannot tell if the same staff member encountered a mystery shopper one or more times. It is true to say that without a control group we cannot truly conclude that it is the intervention that is resulting in the improvement in performance. We believe the number of visits, pharmacies included and the variety of scenarios pooled together demonstrate that the visit number is correlated to both questioning score and outcome. Wording of the conclusion and discussion has been changed to emphasise that we have observed an association between repeated visits and improved outcome and questioning scores.

"The provision of feedback to individual staff, as opposed to the entire pharmacy, means that the study design relies on staff disseminating the information between themselves or the same staff member being shopped multiple times in order for the intervention to be effective."

VERSION 2 – REVIEW

REVIEWER	Bernhard Langer Prof. Dr. Bernhard Langer Professor of Health Care Management University of Applied Sciences Neubrandenburg Faculty of Health, Nursing, Administration Brodaer Straße 2 17033 Neubrandenburg GERMANY fon: +49 (0)395/5693-3107 fax: +49 (0)395/5693-73107 Mail: langer@hs-nb.de
REVIEW RETURNED	06-Nov-2017

GENERAL COMMENTS	It's a real pleasure to read a revision where the authors have so carefully attended to reviewer comments, which made this easy to do. The researchers were made most of my recommendations. Thank you for your effort for these changes.

REVIEWER	Bridget Paravattil Qatar University, Qatar
REVIEW RETURNED	02-Nov-2017

GENERAL COMMENTS	thank you for addressing the reviewer's comments.