Author’s response to reviews

Title: Evaluation of real-world evidence for the effectiveness of academic detailing on appropriate prescribing of pain relief medication in Belgian general practices: a cluster randomized trial

Authors:
Robin Bruyndonckx (robin.bruyndonckx@uhasselt.be)
Veronique Verhoeven (veronique.verhoeven@uantwerpen.be)
Sibyl Anthierens (Sibyl.Anthierens@uantwerpen.be)
Koen Cornelis (koen.cornelis@intermut.be)
Katelijne Ackaert (katelijne.ackaert@cm.be)
Birgit Gielen (birgit.gielen@intermut.be)
Samuel Coenen (samuel.coenen@uantwerpen.be)

Version: 2 Date: 12 Sep 2017

Author’s response to reviews:

Dear Prof. Sevdalis,

Re IMPS-D-17-00254R1

Evaluation of real-world evidence for the effectiveness of academic detailing on appropriate prescribing of pain relief medication in Belgian general practices: a cluster randomized trial

Robin Bruyndonckx; Veronique Verhoeven; Sibyl Anthierens; Koen Cornelis; Katelijne Ackaert; Birgit Gielen; Samuel Coenen

We are grateful for the remarks of the reviewers and would like to thank you for giving us the opportunity to revise the manuscript. We believe we have addressed all comments as itemised in our point-by-point reply below and hope that the responses are satisfactory.
Reviewer reports:

Reviewer #1: Congratulations to the authors for this worthy effort in conducting a pragmatic approach to studying effectiveness of the Academic Detailing Service (ADS), in Belgium. This paper stimulates important reflections on relevant issues about sustainability of the health systems as well as pharmaceutical industry, based on rational use of drug resources and balancing between commercial and marketing purposes. Assessing the effectiveness of ADS on the application of the recommended protocol for pain relief in osteoarthritis, throughout the reimbursements of analgesics and NSAID prescriptions, the paper provides relevant information that support a successful public policy in the best interest of the population who is covered by that, in Belgium, and other countries that have similar ADS.

> We thank reviewer 1 for taking the time to go through the manuscript so thoroughly, appreciate these positive words and value his suggestions.

Abstract

The main scientific purpose is not stated, instead only that it was ordered by the Federal Agency for Medicine and Health Products (FAMHP) to Farmaka ADS. This alone is not enough to justify the study.

> We clarified the main purpose of the study by rewriting the background section of the abstract in lines 29-32 as follows:

In Belgium, the debate about the effect of the national academic detailing service (ADS) on prescribing quality in general practice is ongoing. In order to evaluate both the implementation strategies of the ADS and its effectiveness on appropriate prescribing of pain relief medication, we conducted a real-world cluster randomized controlled trial (cRCT).

The method is quite complex related the sampling process, types of eligible practices, and GP coding. There are too many details that are difficult to understand. I suggest making the methods more intelligible for the readers with a more comprehensive and synthesized description.

> We simplified the methods section in the abstract by removing details on eligibility and randomization.

The results also have too many details about the randomization process and the final analyzed sample. I suggest reporting the main results about the effectiveness and impact of the academic detailers.
Introduction

The first part raises the problem comparing published data on the effectiveness of Continuing Medical Education (CME) and ADS. The authors cite a Cochrane's review that attributes small improvements in drug prescribing related to ADS. A recent synthesis of systematic reviews about CME impact published by Cervero & Gaines, has shown significant but small effects on physician performance and patient health outcomes, of 6% and 3%, respectively (Cervero RM, Gaines JK. The impact of CME on physician performance and patient health outcomes: an updated synthesis of systematic reviews. J Contin Educ Health Prof. 2015;35:131-8.) I think that this information can be relevant to the readers make judgments about the effectiveness of both different strategies.

> We thank reviewer 1 for bringing this interesting review paper to our attention and included a citation in lines 68-70.

The second paragraph promotes awareness about the FAMPH request of this effectiveness trial based on insufficient evidences published in two previous small RCT studies, although they have shown effectiveness in antibiotic and benzodiazepines prescribing. Thus, I wonder that if this request is an updating research process or it was triggered by any other reason.

> The analysis of the effectiveness of the ADS was ordered because FAMHP nearly doubled its yearly investment in its ADS in 2009. It was initially investigated by the Belgian health care knowledge centre (KCE) on visits covering dementia and diabetes, but as this study was too small to draw valid conclusions, they reordered it (see lines 78-82).

Moreover, as I mentioned in the Abstract's comments, would be helpful to know why the authors chose the appropriate use of pain relief drugs for chronic pain in osteoarthritis (OA), to be evidenced.

> We chose to study pain relief medication following recommendations KCE made. They suggested using a common condition in which the GP plays a major role and for which an impact could be expected (see lines 90-94).
In the third paragraph, it is stated that OA is a common condition that NSAIDs are often prescribed increasing risks of gastrointestinal complications. The references list brings 7 articles about OA and NSAID use bleeding complications, but just one is cited in the introduction (#10). As a reader, I think there is a need for a more detailed approach of this issue in Belgium, handled by health authorities and Farmaka, in my opinion.

> We included additional references to provide more detailed background information for interested readers (references 11-17). The reason FARMAKA zoomed in on pain relief medication for OA is because OA is a common condition that is occurring more and more frequently due to the ageing population and for which a large amount of non-steroidal anti-inflammatory drugs are prescribed, regardless of the associated side effects (see lines 95-97).

Methods

It is useful that this section comes splitted in sub-sections. It is understandable that the randomization was performed over the practices instead of individuals GPs to minimize contamination bias between GPs, in group or duo practices. I cannot understand what mean these different practices and how it improves the data reliability. The authors explain that the practices were performed in permuted blocks of two (exposed x control group, I suppose), and stratified according detailer, visits, and type of practice as well as they were sorted by GP's descending code. This explanation needs more clarification to whom is not familiarized with these practices and GPs, in Belgium context. It is not a technical questioning about the method used, but what are the practices and their types, and how GP's codes identify practices. There is no mention about them in the whole paper. On the other hand, the intervention is well described as well as the Key Messages (KM).

> We are pleased that the key messages came across clearly, but (unfortunately) did not realize that non-Belgian readers might not be familiar with the Belgian system of GPs within practices. We appreciate reviewer 1 pointing this out to us, and clarified that the practice types include practices with one, two and more than two GPs. We also indicated that GPs get their identification code from the Government, instead of from the research group, and that a higher ID implies a longer time since registration (see lines 117-121).

The data characteristics sub-section provides the information that the practices, GPs, and Academic Detailers data, were provided by Farmaka, and the additional important data were provided by the Intermutualistic Agency. There is no mention about the reliability of this dataset, although the authors remark that all data had a unique identification code. Thus, how these datasets were built or joined, seems important to be mentioned.
We provide more information on the data collected and provided by Farmaka and the Intermutualistic Agency, and clarified that both datasets were linked using the unique GP identification code which is provided by the Belgian government upon registration of a GP and is available in both datasets (lines 120-121 and 169-171).

The primary and secondary outcomes are well explained and the complex process of intention-to-treat analyses in different scenarios are well described and depicted in figure 3, as well as the statistical modeling.

Results are well described and depicted in the tables and figures. It is not so easy to understand them in function of the multiple scenarios modeled for ITT and per protocol analyses.

We are glad that the outcomes and different scenarios were clearly understood. The four different scenarios were included to assess the robustness of our findings. This was clarified by including lines 290-291.

Discussion

The discussion misses important issues that could be related to these results. There is no mention about the main problem raised by the study question, which is the effectiveness of ADS on chronic pain relief drug prescribing by GPs.

We believe the effectiveness of the ADS on chronic pain relief drug prescribing was discussed in lines 436-443.

‘In this study, we assessed the effectiveness of Farmaka’s AD visits on appropriate prescribing of pain relief medication in osteoarthritis. We found a significant impact of the visits conducted by academic detailers on one primary outcome i.e. the proportion of patients prescribed with a recommended NSAID among those prescribed with any NSAID, confirming the usefulness of the ADS in improving appropriate prescribing of pain relief medication.’

Although the authors reinforce their findings agreeing with the former RCTs about effectiveness of antibiotics and benzodiazepines, in Belgium, would be interesting to have more information about similar studies of other countries about this issue, mainly when they pointed out that the impact of ADS might be very condition and outcome specific. Is there any information in the literature reinforcing or not this hypothesis? The Borgermans et al. paper cited is written in Flemish that restricts this knowledge to non-Flemish speakers. I think that it needs stronger evidences to support this statement.
Firstly, we apologize for including the report from Borgermans et al. in Dutch, as it is also available in English. We changed the reference to the English version such that also non-Dutch speakers can consult this report.

We also added some references to research on the impact of ADS in other countries, where some authors did find impact (references 25 and 26) and others did not (reference 27). We hope that this provides sufficient evidence to support our words.

The other significant findings related to Academic Detailer profile is low explored by the authors, as well. There are no information about them, who could appear in the Introduction. It is no possible to wonder why being older or physician could promote higher impact in drug prescription.

Because the impact related to Academic Detailer profiles were studied in an exploratory analysis, for which this study was not initially powered, we decided not to draw too much attention to these findings. Nevertheless, we elaborated on them a bit more in detail in lines 454-460.

The sub-section "On the use of reimbursement data" pointing out missed information about Paracetamol, seems to me one major study limitation that was neither mentioned at the Methods, nor as a limitation. The author's strategy to analyse these data only in patients 60 and older, and that they get paracetamol reimbursement only by prescription, could improve the reliability about the significant small impact on paracetamol dosage finding. Moreover, the lack of consensus on the paracetamol recommendation in OA, is stated by the authors as a major study limitation. I do not agree, since the main result was related to the recommended NSAID adequacy prescription, although it can be a really limitation.

We included the section on the use of reimbursement data as a limitation (lines 485-495) and specified the limitation of the lack of consensus on the paracetamol recommendation as a limitation to PO2 and SO2 (lines 507-509).

The other limitations, i.e., missing by almost a quarter of eligible practices and short-term effects should be more clarified to give the readers the dimension of how these limitations could impact the conclusions.

Finally, the pragmatic cluster RCT design has brought robustness to the findings, aside the large sample of practices analysed.
Because we focussed on findings that were consistent over the four scenarios, and these GPs were included in the ITT scenario, the 24% unreached practices will have no implications on the findings (see lines 514-520). Additionally, we clarified that the 6-month time frame implies that the sustainability of effects more than six months after the AD visit could not be assessed (lines 530-531).

Conclusions

The authors conclude that only impact of the study was the improvement in NSAID adequacy prescription, but there were other impacts related to detailers and paracetamol prescribing that could be added.

Because the impacts related to detailers were studied in an exploratory analysis, for which this study was not initially powered, we decided not to focus on these findings. We mentioned the impact of the detailers’ characteristics that could be of importance (i.e. age and training) but decided not to discuss this extensively and to detail those findings in the conclusions.

Reviewer #2: This is a cluster randomised trial to evaluate the effectiveness of an academic detailing visit by Farmaka ADS on prescribing of pain relief medication. This is a reasonably well-written paper, although the presentation might be improved by a clearer structure (lists, headings etc.) instead of long text paragraphs.

We thank reviewer 2 for taking the time to go through the manuscript so thoroughly and value her suggestions. In order to improve the structure of the paper, subheadings were added to the methods section (lines 268, 332 and 346) and section Outcome definition was restructured (lines 182-232).

The intervention was allocated at the GP practice level (cluster). Summary patient outcomes were also determined at the GP level.

The information for this study was delivered at GP level. However, in our analyses, all this information was aggregated at the level of the practice.

My main criticism regarding the reporting of this cluster randomised trial is that the primary outcome (or outcomes) have not been defined. Instead a time series of monthly pre- and post-
intervention measures is modelled using a fairly complex approach and various parameters are tested, but without clarification as to why these parameters would provide evidence of effectiveness.

> The primary (and secondary) outcomes were defined in the section Outcome definition, but we have now restructured this section in order to provide more clarity hereon (lines 182-232). The rationale for choosing these outcomes was added to this section.

Detailed comments:

Trial design:

3529 practices were randomised but only 1698+1703 practices were included in the analyses. I assume this is because "GPs were given the option to opt out from the analyses". This seems an unfortunate design choice/consent procedure. All the problems associated with non-random treatment allocation are potentially re-introduced by creating non-response informed by treatment allocation. Why were GPs who were not willing to take part in the study not excluded before randomisation to safeguard the internal validity of the trial?

> The 128 practices (196 GPs) were excluded for varying reasons (listed in Figure 1):

  • because we could not assess whether they had been visited or not (lost to follow-up):
    0 (5) in intervention group and 0 (6) in control group.
  • because they had duplicate records in the database. Rather than choosing one or the other, we decided to remove both entries:
    8 (13) in intervention group and 6 (14) in control group.
  • because they chose to opt out (most likely these GPs stopped practicing):
    14 (20) in intervention group and 21 (27) in control group.
  • because their GP ID could not be linked with prescribing information on analgesics (e.g. were inactive at the time of collecting this information):
    31 (48) in intervention group and 16 (28) in control group.
  • because we could not assess the number of patients that visit the practice yearly:
    13 (15) in intervention group and 19 (20) in control group.

This implies that only one in four GPs was excluded because they chose to opt-out, which limits the impact on the analyses.
We chose to provide the option to opt-out at the latest possible moment in time, right before analysing the data collected and provided by Farmaka and IMA. This was done in order to be able to provide real-life evidence of the impact of the visit on the GPs’ prescribing behaviour, regardless of the study. When made aware of the study from the start, GPs might be inclined to alter their prescribing behaviour for the duration of the study in order to ‘look good’.

Why were the practices sorted by age of GP before randomisation? Are the blocks of two to represent age strata? You mention creating "a comparable spread of visit dates" but the control group does not receive any visits?

> Sorting of the practices by 'time since registration of the GP registered first’, within strata, was done to optimize comparability of practices in the permutation blocks. This way, we avoid matching of practices with a more experienced senior GP with a recently graduated senior GP (e.g. in a solo practice). The blocks of two don’t represent age strata, but the earlier blocks contain practices with a more experienced senior GP compared to the later blocks. These permutation blocks don’t carry any information, but were used in order to optimally randomize the eligible practices. This was clarified in line 118.

The control group does not receive a visit, but because within each block each control practice was matched with an intervention practice during randomization, they would have received a visit on the date the intervention practice was visited, and therefore, this date is carried over from the intervention practice to the matched control practice (lines 122-125). We need a comparable spread of visit dates because we include prescribing information 6 months before and 6 months after this date.

Methods:

The presentation might benefit from a list of measures (bullets, table etc.)

> A list of measures was presented in Figure 2. We believe that including an additional table here would be redundant to the information in this figure.

Analysis approach/Figure 2:

This section would benefit from a clearer description of the research questions asked and the approaches used to address them.

> The research question for each outcome is whether or not the key message (depicted in Figure 2) was conveyed. Because it might not immediately be clear in which direction the outcomes
should evolve, we indicated what the four primary and four secondary outcomes are, how they are calculated, and what we expect if the key messages were delivered successfully (which equals the alternative hypothesis) (see lines 182-232).

The analyses labelled ITT1 and ITT2 are not intention to treat analyses as not all observed subjects are analysed in the groups to which they were randomised. I would consider all of ITT1, IIT2 and PPR per-protocol approaches based on different definitions as to what constitutes a protocol violation.

> We adjusted the definition of ITT1 and ITT2 to intention to treat with violation (ITT-V1 and ITT-V2, respectively) and adapted Figure 3 accordingly.

I would assume that the analyses are carried out for different purposes? If their purpose is to also evaluate efficacy (=the effect of receiving a detailing visit) there might be better ways of achieving that.

> The purpose of the analysis of different scenarios was to assess the robustness of the findings from the ITT0 (now ITT) analysis. This is clarified in lines 290-295.

Statistical methods:

It is not clear whether repeated measures/multilevel modelling or time-series methods are used here and how they are motivated. Time-varying explanatory variables cannot be the motivation as trials do not typically require adjustment for baseline variables. There is some mention of using a "working correlation matrix" which hints at the use of GEE (presumably with robust standard errors)?

> Both repeated measures and interrupted time-series models are used. The motivation for this:

• We wanted to allow a gradual change (trend), a sudden shift at the moment of the intervention and change in trend after the intervention. For this, we used an interrupted time-series model.

• Because practices were clustered within province (Belgium has 10), and we expect practices within the same province (and hence the same academic detailer) to display more similar prescribing behaviour compared to practices from a different province, we corrected the analysis for clustering of the measurements within province, which turns the model into a GEE.
• Because the models in the explanatory subgroup analyses include covariates which are not randomized by and could vary with time (the study covers years 2012 and 2013 and e.g. the number of patients visiting the practice yearly changes yearly), we used an independent working correlation.

• Because model-based standard errors don’t provide correct adjustment for repeated measures, and providing both model-based and robust standard errors would bring confusion, we decided to report only robust standard errors without mentioning this explicitly.

We clarified the fact that the interrupted time series model was adjusted for clustering within province. The use of time-varying covariates was clarified and the use of standard errors was explicitly mentioned (lines 300-308).

The analysis seems overly complex and does not target a clearly defined primary outcome. Which of the 6 model parameters represent measures of effectiveness and are therefore of interest here? I would expect the objective of a clinical trial analysis to be demonstrating that an intervention has an effect on a primary outcome measured at a specified post-randomisation time point.

> Because we expected the impact to be gradual and wanted to include the possibility for an effect to play e.g. at the first month after the intervention which would then fade out again, we used this more complicated model. The parameters of interest therefore are β6 and β7. The first represents the sudden change caused by the intervention at the moment of the intervention (i.e. the step change). The latter represents the change over time following the intervention, indicating whether the effect of the intervention fades out or sustains (i.e. the change in trend). This is briefly explained in lines 324-327, and a cross-reference to the naming in the tables in the results section is included to improve readability.

There are a number of subgroup analyses. I am not sure that these are necessary. In addition, any modification assessment would need to be modelled by various interaction terms.

> The subgroup analyses were included because they provide interesting information on the most successful implementation strategies of the visits which will be interesting for the readers of Implementation Science. Assessment of the impact of additional covariates was indeed done by including several two- and three-way interaction terms with the additional covariate.

How does the analysis deal with missing data? What assumptions are made?
Because we excluded practices for which information on analgesics consumption or information on the number of patients visiting the practice yearly was missing, this is a complete case analysis.

In the subgroup analyses, the information on the age and training of the academic detailer could be missing. Because missingness proportions were maximally 7% of studied practices, these were excluded from the corresponding subgroup analyses (and only from these subgroup analyses, they were included in the other subgroup analyses). Also this therefore corresponds to a complete cases analysis. Imputing information when the percentage of missingness is this small and subgroup analyses are merely exploratory seems to be overcomplicating things.

A number of baseline variables were used in the randomisation procedure. Was this acknowledged in the analysis approach?

The statistical analysis did not include baseline covariates, because the practices were randomly allocated to either intervention or control group based on these baseline variables. Therefore, any observed difference between the intervention and control group must be purely by chance (which is what a significance test then would test for).

Results:

This section implies that the estimates of interest are step change = difference before and after the intervention and change in trend = monthly change after the intervention (see table 2). I would be looking for comparisons between trial arms. I was not able to understand the meaning of the effects quoted in table 2.

We thank the reviewer for this remark, as the naming in the table is indeed confusing. It are parameters $\beta_6$ and $\beta_7$ we are interested in, and which are therefore represented in Tables 2 and 3. We clarified this by including a reference to the column names that are used in Tables 2 and 3 in the text (in lines 324-327), including a reference to the parameters mentioned in section Statistical analysis in Tables 2 and 3, and by expanding the explanation for the column names underneath Tables 2 and 3. We hope this clarifies things sufficiently.

The results section discusses change over time, which would not be relevant in the context of establishing effectiveness.

We were interested in the change over time because we wanted to know whether the effect of the intervention (if found) would last, rise or fade after the intervention itself.
If the effects quoted in tables 2 and 3 for proportion outcomes represent odds ratios then I would be concerned about their clinical relevance. Two significant effects seem to indicate a reduction of 2%?

> The effects presented in Tables 2 and 3 are indeed odds ratios (except for the effects for PO2 and SO1, which are average DPM). As you mention, there are two effects which are not of clinical relevance. For completeness, we did include a mentioning of their statistical significance (p-value < 0.05). But since these effects were not significant in the four studied scenarios (i.e. not underlined), and we focussed on effects that had consistent significance over the four scenarios, we did not report on these two effects any further.

Even though the primary outcomes have not been clearly specified, there are at least four (more if several months are of interest). No adjustment for such multiple comparisons has been applied.

> You are indeed correct that this should have been done for the four primary outcomes. Therefore we adjusted the significance threshold to 0.01 (instead of 0.05/4 = 0.0125 for convenience) and the confidence intervals to 99% CI instead of 95% CI in Tables 2 and 3.

Tables 4 and 5 labelled "covariate effects..." are simply provided and not referred to in any detail. I was unable to understand their purpose.

> Tables 4 and 5 were referred to in line 410, with an indication of location to appear in printed text on lines 415 and 417. These tables represent the results of the exploratory subgroup analyses, which are discussed in section exploratory subgroup analysis starting on line 409.

--

We believe that this revised manuscript is better suited for publication in Implementation Science as compared to the original version and hope you are willing to reconsider this version for publication.

With kind regards,

Robin Bruyndonckx,
on behalf of all co-authors.