

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A six-year comparative economic evaluation of healthcare costs and mortality rates of Dutch patients from conventional and CAM GPs
AUTHORS	Baars, Erik; Kooreman, Peter

VERSION 1 - REVIEW

REVIEWER	Helge Bruns Department of General and Transplant Surgery, University Hospital Heidelberg, Germany
REVIEW RETURNED	03-May-2014

GENERAL COMMENTS	<p>Baars et al. present an interesting retrospective analysis on economic impact of complementary and alternative medicine (CAM) performed by general practitioners (GPs) on Dutch patients. The authors have used data from an insurance company (Agis) and have compared annual costs of patients treated by GPs without to patients treated by GPs with training in CAM. The authors identified lower annual costs for those patients while there was no difference in mortality.</p> <p>Unfortunately, there are some limitations that need to be addressed:</p> <p>It is not clear whether the two groups of patients can be compared; there might be a difference in both patient collectives. It might be that patients with severe diseases tend to avoid CAM (or vice versa), or that GPs avoid CAM in some patients based on personal preferences or depending on the underlying disease (the authors state that some of the GPs practised both CAM and no CAM treatments). This is an important point that needs to be discussed; the authors need to compare data from both patient collectives for homogeneity of age, underlying diseases etc. The most appropriate way to avoid this kind of bias would be to perform the study as a matched pair analysis.</p> <p>It is not clear whether a decreased annual cost of 192 Eur is related to decreased treatment costs or to reimbursement; since the authors have used data from an insurance company, it could be that patients who were treated by GPs with CAM had to pay part of their GPs bill themselves. Costs provided by insurance companies (=reimbursement) are not necessarily related to real costs of patient treatment.</p> <p>The authors need to specify "CAM"; they simply state that the "majority of the CAM GPs are anthroposophic GPs (70 AM practices (64%)). What did the other 36% of CAM GPs do? What kind of diseases were treated using</p>
-------------------------	--

	<p>CAM/no CAM?</p> <p>As far as statistics are concerned, the authors admit they have used t-testing. Most likely, data on costs is not normally distributed and thus t-testing might not be appropriate.</p> <p>Taken together, this is an interesting study, but due to the limitations mentioned above, the conclusions drawn by the authors can hardly be justified. The reviewer suggests to perform this study as a matched pair analysis, which would lead to robust findings.</p>
--	---

REVIEWER	Patricia M Herman The RAND Corporation USA
REVIEW RETURNED	10-Jun-2014

GENERAL COMMENTS	<p>My 'yes' answers to #s 3, 4, 7, 9, 10, 11, 12 and 15 above should be tempered by my comments below.</p> <p>Study sample - can you give more information about who was included in the sample? For example, did the whole sample have both compulsory and supplemental insurance? If not, is having supplemental insurance a matter of being able to afford it? Do all supplemental insurance plans cover CAM and what kinds do they cover? The use of claims data already limits us in terms what we can know about this sample, but these data contain more information than is presented.</p> <p>In the model overview (page 8) there is no explanation as to why particular methods will be used. What is your reasoning behind your choices of methods? Also, clearly lay out the variables you had available in the data set to match/adjust groups and the reasoning behind them.</p> <p>Page 9 you say that you exclude "switchers." This is the most interesting group! These are the only people who had both CON and CAM. I'd like to know how many switched from CON to CAM and how many from CAM to CON. Were the percent switchers (compared to all CAM or CON) the same in both cases or did people tend to switch one way more than the other? And most important, how did their annual costs change after they switched?</p> <p>Another analysis which would add to the evidence that there is a true difference here (and not just evidence of selection bias) is to take a subset of each group (CON and CAM) which have some similar diagnosis, if you can find a group where the CAM sample is still big enough. And/or if CAM is only available in certain areas (and only in small areas), you could compare the costs of all those (CAM and CON) in the areas where CAM is available to all those in areas where CAM is not available to get an idea of the impact of CAM availability.</p> <p>I don't see anywhere how you dealt with the fact that you had 6 years of data. Were yearly numbers averaged? Was any adjustment for inflation made? Was there any attempt to look at trends?</p> <p>You spend a whole table (Table 3) giving the results of the log linear model when you could simply say something like 'the results of the</p>
-------------------------	--

	<p>log linear model for the most part replicated/supported the results of the linear regression model.' Please resist the impulse to use a log linear model because of the skewedness of the cost data. The coefficients do not mean anything and with large sample sizes straight linear regression has been shown to be sufficient. Also, at the bottom of page 11, you do not need to repeat in text all the information in the Table 2. Mainly save your text for key points you want people to get from the table and especially waste no text talking about $p > .05$ but less than .1 results. Similarly, in the discussion section starting on page 14, you don't need to repeat all the detailed statistics.</p> <p>Page 16, lines 44-45, instead of giving a 0-30% range for the results of your previous study, can you give the number that most closely parallels the 12/4% found here?</p> <p>Page 19 - the manuscript abruptly ends with a ; Was there more text that I did not see?</p> <p>Minor edits</p> <p>Page 3 line 17 - 'prohibiting the possibility' is maybe too strong; suggest something like 'reducing the ability' There are methods other than RCT which can be used to give indications of causality</p> <p>Page 6 lines 8-15 - since this is a quote you need a reference</p> <p>Page 6 line 43 - I assume you are referring to your previous study, so maybe "an initial" economic evaluation. Using "a first" somewhat sounds like it might be the first step/part of this study.</p> <p>Page 8 line 34 - is "Agis insures" supposed to be "Agis insurees" or "Agis insurers"?</p> <p>Page 10 Table 1, put the "n's" at the top of each column.</p> <p>Page 13 line 54 - I think you mean high p-values, not low ones</p> <p>Page 14 Table 4, what are the units?</p> <p>Page 14 line 42 maybe instead of 'but are the highest' when you are talking about a negative cost, use 'but the largest savings'</p> <p>Overall a good paper. It just falls short of what it could be.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer Helge Bruns

- Are the two groups comparable?:
 - o **Our response:**
 - As described in the Methods section (*Dataset on healthcare costs and demographics*) the dataset does not contain any information on health status besides mortality. We therefore were not able to control for possible differences in health status. We controlled for all relevant variables that were available: age, sex, year, and socio-economic status (4-digit postcode).
 - In the Discussion section we discussed the possible differences in health status between the groups and summarized the existing literature on this topic: *“With respect to selection, several studies that compare the health status of patients treated in CAM and in conventional medicine in primary care settings find that patients treated in CAM practices suffer more often from severe and chronic illnesses (e.g., [16, 17]). This suggests that if we could control for severity and chronicity of illnesses (with additional data), the estimated compulsory cost differences might be larger.”*

- A matched pair analysis based on the available variables would not solve the problem of unobserved differences between the two groups of patients.
 - **Article changes:**
 - We did not change the text, since with the limitations of this database, we are not able to perform analyses on other variables (e.g., disease state) and therefore other statistical analyses (e.g., matched pair analysis) are not indicated. Since we have described these limitations sufficiently in both the Methods and the Discussion sections, we did not change the text.
- Treatment cost and reimbursement:
 - **Our response:**
 - All Dutch people have a compulsory insurance that covers most of the costs of GP care, pharmaceutical care, hospital care and some paramedic care and that are paid for by the insurance companies. Additional care (e.g., specific or more paramedic care, CAM care) must be paid for by the individual out-of-pocket and/ or can be covered (up to a maximum) by the supplementary insurance (e.g., costs of CAM treatment is paid for up to 500 Euro/ year). We changed the text to make this more clear.
 - In the Discussion we discuss this topic: *“Fourthly, the lower costs could be related to the fact that patients interested in CAM might have higher out-of-pocket expenses since not all CAM treatments are covered by supplementary insurance. Clarifying the role of out-of-pocket expenses is an empirical issue that requires additional data.”*
 - **Article changes:**
 - We changed the text *The Dutch financing system* in the Introduction to make the costs covered by the compulsory insurance more clear: *“The Dutch financing system contains two basic compulsory health insurances that are for 80% paid for through income taxes: for curative care (Zorgverzekeringswet (ZvW)) and for long-term care (Algemene Wet Bijzondere Ziektekosten (AWBZ)). The compulsory health insurances cover costs of most of GP, pharmaceutical and hospital care and some paramedic care. In addition, people in The Netherlands can buy supplementary insurance. Supplementary insurance covers costs not covered by basic insurance (for example specific or additional paramedic treatment, complementary therapies) (e.g., costs of CAM treatment is paid for up to 500 Euro/ year) [5]. Supplementary insurance can also cover the costs of improvements over the standard level of care paid for by compulsory insurance (e.g., extra costs for a better room and service in case of hospitalisation).”*
- CAM modalities:
 - **Our response:**
 - We agree that it is more transparent to describe the other CAM modalities also. We changed this in the text.
 - **Article changes:**
 - We added to the text *Results - GP practices and patients: Other CAM GPs were specialized in acupuncture (15%) and homeopathy (25%). Since some GPs were specialized in more than one CAM modality the total percentage of CAM GPs is larger than 100%. Exact numbers and percentages of CAM GPs vary a little over the years.*
- Kind of diseases treated by CAM/non-CAM:
 - **Our response:**
 - See the reaction to the first item *Are the two groups comparable?*
 - **Article changes:**

- We did not change the text, since we have described this topic sufficiently in both the Introduction and the Discussion.
- T-testing:
 - **Our response:**
 - The reviewer is right in pointing out that costs are unlikely to be normally distributed, even after taking logs. Given the large sample sizes available here, asymptotic t-testing for differences in means is appropriate by virtue of the central limit theorem. Applying non-parametric procedures is a topic for future research.
 - **Article changes:**
 - We added to the text *Methods – Statistical analyses: Given the large sample sizes available here, asymptotic t-testing for differences in means is appropriate by virtue of the central limit theorem.*
- Conclusions can hardly be justified:
 - **Our response:**
 - We do not agree with the reviewer. The conclusions are based on the comparison we made between the groups in this dataset; in other words, the conclusions pertain to this specific study. In the Discussion section we have thoroughly discussed the limitations of the study and their influence on the generalizability of the results of this study. We also discussed possible explanations for the results, and described main future research lines that should provide answers to the remaining questions.
 - **Article changes:**
 - We did not change the text, since we have described this topic sufficiently in the Discussion.

Reviewer Patricia Herman

- Study sample:
 - **Our response:**
 - As described in the Introduction compulsory health insurance is mandatory for all patients in The Netherlands. So there is no difference with regard to this between the two groups. The percentages of patients that have a supplementary insurance are presented in Table 1, but this was not described in the text. Many Dutch supplementary insurances cover costs from complementary treatments, especially the three CAM modalities that are studied here: anthroposophic medicine, acupuncture and homeopathy. We changed the text with regard to these last two topics.
 - **Article changes:**
 - We added text to *Results – Healthcare costs – The dataset: The percentages of patients with a supplementary insurance were almost the same (CON GPs: 92.7%; CAM GPs: 93.4% and Switchers: 92.1%).*
 - We added text to *Introduction – The Dutch financing system: Many supplementary insurances cover costs of CAM treatments like anthroposophic medicine, acupuncture and homeopathy.*
- Model:
 - **Our response:**
 - We use a regression approach to control for observed differences in patient characteristics: age categories, gender, dummy variables 4-digit postcode, etc. [see p.7 in the article]. These variables are correlated with both costs (and mortality) as well as with the likelihood of choosing a CAM GP, and

therefore important to control for to mitigate confounding as much as possible. The regression approach is standard practice in health economics and yields results similar to those of matching procedures (both are unable to correct for unobserved differences between groups of patients). Given the large sample sizes Students t tests are asymptotically valid by virtue of the central limit theorem, independent of whether the underlying distributions are normal or non-normal. Standard errors are clustered at the level of the insured to control for the statistical dependence of observations pertaining to a given insured person (i.e. observations are independent “between” individuals but dependent “within” individuals).

- **Article changes:**

- We added text to *Methods – Model overview*: *The regression approach is standard practice in health economics and yields results similar to those of matching procedures (both are unable to correct for unobserved differences between groups of patients). Given the large sample sizes Students’ t tests are asymptotically valid by virtue of the central limit theorem, independent of whether the underlying distributions are normal or non-normal. Standard errors are clustered at the level of the insured to control for the statistical dependence of observations pertaining to a given insured person (i.e. observations are independent ‘between’ individuals but dependent ‘within’ individuals).*

- Switchers:

- **Our response:**

- We agree that it is good to describe these aspects. We changed this in the text.

- **Article changes:**

- We added the following text for the Appendix section:

The Switcher group

From the total group of 10,769 Switchers, during the period 2006-2011, 6,224 patients switched one time; 2,992 patients switched two times (= back to their first type of GP); 1,282 patients switched three times; 241 patients switched four times and 30 patients switched five times. From the Switchers group that started with a CAM GP, 69.3% ends up with a CON GP. From the Switchers group that started with a CON GP, 70.5% ends up with a CAM GP. As a result the total percentages of CAM patients and CON patients hardly change.

When we analyze the changes in compulsory costs after switching in the subgroup that switched only one time, the total compulsory costs after switching are higher, independent of the direction of the switch. Switching from a CON to a CAM GP results in 337 Euros higher costs ($p < 0.001$), switching from a CAM to a CON GP results in 372 Euros higher costs ($p < 0.001$). After correction for observed differences between the groups by means of linear regression analyses, switching from a CON to a CAM GP results in 34 Euros lower costs (not significant: $p = 0.83$) and switching from a CAM to a CON GP results in 360 Euros higher costs ($p < 0.079$).

When we analyze the changes in supplementary costs after switching in the subgroup that switched only one time, we see that switching from a CON to a CAM GP results in 23 Euros higher costs ($p < 0.001$), and that switching from a CAM to a CON GP results in 1 Euro lower costs (not significant: $p = 0.78$). After correction for observed differences between the groups by means of linear regression analyses, switching from a CON to a CAM GP results in 1 Euro higher costs (not significant: $p = 0.816$) and

switching from a CAM to a CON GP results in 2 Euros higher costs (not significant: $p=0.803$).

Since we are mainly interested in the differences in costs between patients that have a CAM GP and patient that have a CON GP for the whole period of six years (2006-2011), the Switcher group is left out of the following analyses.

- Subset with similar diagnoses:
 - o **Our response:**
 - See the reaction to the first item *Are the two groups comparable?* (reviewer Helge Bruns)
 - o **Article changes:**
 - We did not change the text, since we have described this topic sufficiently in both the Introduction and the Discussion.
- Six years of data:
 - o **Our response:**
 - We used the data set as a panel. This means that if an insured person is observed for all six years, we use six observations of annual costs of this person in the analysis (taking into account the “within”-person correlation by clustering standard errors at the level of the individual). The reported differences can be interpreted as the average of cost differences across years. Any trends are controlled for by the year dummy variables. However, trends are not the primary focus of the analysis. Costs were not corrected for inflation. Since inflation was low (about 2% during 2006-2011), correcting for inflation is unlikely to affect the key results.
 - o **Article changes:**
 - We added the text to *Model overview*: *With regard to the six years of data the data set was used as a panel. This means that if an insured person is observed for all six years, six observations of annual costs of this person are used in the analysis (taking into account the ‘within’-person correlation by clustering standard errors at the level of the individual). The reported differences can be interpreted as the average of cost differences across years. Any trends are controlled for by the year dummy variables.*
- Table loglinear model:
 - o **Our response:**
 - In line with the presentation of the results in our first article and in line with the standard in (health) economic journals, we think it is important to present here also the results of the log linear analyses. We regard it important to describe the differences in results from the linear and the log linear analyses. However, in line with the request of the reviewer we will shift the table to an online appendix.
 - o **Article changes:**
 - We shift the table from the article to an online appendix.
- Previous study: range 0-30%:
 - o **Our response:**
 - In our first study we did not calculate the average annual total compulsory healthcare costs of patients treated by a CAM GP for the total group, we just analysed the cost differences for the specific age groups. In addition, in the first study, we were not able to distinguish between compulsory and supplementary costs. Therefore, we are not able to present “the number that most closely parallels the 12.4% found here”.
 - o **Article changes:**
 - We did not change the article, since we do not have the requested results.

- Ending of the manuscript on page 19:
 - o **Our response:**
 - We don't know why you did not have more than 19 pages, since we have submitted the whole concept article.
- Minor edits:
 - o P. 3 line 17: we changed 'prohibiting the possibility' into 'reducing the ability'
 - o P. 6 lines 8 – 15: we added a reference
 - o P. 6 line 43: we changed 'a first' into 'an initial'
 - o P. 8 line 34: we changed 'insures' into 'insurees'
 - o P. 10 Table 1: we put the N's in the first row of the table
 - o P. 13 line 54: we changed 'low' into 'high'
 - o P. 14 Table 4 (units): In the LPM model the coefficient (first column) means that the probability of dying in the six year period for CAM patients is 0.4% lower than for CON patients. (The base level probability is about 5.2%) Logit coefficients are not readily interpretable. Additional calculations show that according to the logit model the probability of dying in the six year period for CAM patients is about 0.3% higher than for CON patients. In the Cox proportional hazard model, the first coefficient implies that the hazard rate (the probability of dying in an infinitesimally small interval, conditional on not dying before) is 5.9% higher for CAM patients.
 - o P 14 line 42: we changed 'but are the highest' into 'but the differences are largest'.

We hereby send you the revised manuscript, where all the revised text parts are visible. We thank you again for considering our work for possible publication in the *BMJ Open*.

VERSION 2 – REVIEW

REVIEWER	Helge Bruns University Hospital Heidelberg, Department of Surgery, Heidelberg, Germany
REVIEW RETURNED	29-Jul-2014

GENERAL COMMENTS	The authors have adequately responded to both reviewers' criticism and the manuscript is now acceptable for publication.
-------------------------	--

REVIEWER	Patricia Herman RAND Corporation USA
REVIEW RETURNED	26-Jul-2014

GENERAL COMMENTS	Overall, the authors have responded sufficiently to the comments made. I especially appreciate the analysis of the Switcher group. It is interesting that for compulsory costs the results for the Switcher group validate your overall results. In your overall results the CAM GP group had 225 Euros lower compulsory costs. In the Switcher analysis switching from a CON GP to a CAM GP lowered compulsory costs and switching from a CAM GP to a CON GP increased costs. MINOR EDITS Page 3, 3rd bullet, the word "treatment" should be plural
-------------------------	--

	<p>Page 9, end of the "General practitioners and patients" section, the text says that the Switchers group "were excluded from all analyses." However, they are addressed in the text on the next page, in Table 1 and now in the Appendix. The last sentence of Appendix 1 says that the Switcher group "is left out of the following analyses." But there are no following analyses.</p> <p>Page 12, 2nd full paragraph, refers to a Table 3 that no longer exists.</p> <p>Page 14, new Table 3, please state the units, for at least the LPM model results.</p> <p>Page 20, the article ends with a long sentence with four recommendations for future research. Since there is no "and" before #4 and since it ends with a semicolon (;) rather than a period (.) it looks as if there is text missing.</p>
--	---

VERSION 2 – AUTHOR RESPONSE

With regard to the minor edits of the first reviewer: we have processed all comments (visible with track changes) but one ("Page 14, new Table 3, please state the units, for at least the LPM model results."). Since we have described in the Methods section (subsection 'Model overview') that these are regression models, we think it is clear enough that in Table 3 regression coefficients are presented. It is now presented in the same way as Appendix 2 (also without describing that these are regression coefficients).