PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A study protocol for a quasi-experimental claims-based study evaluating ten-year results of the population-based integrated health care model 'Gesundes Kinzigtal' (Healthy Kinzigtal): the INTEGRAL study.
AUTHORS	Schubert, Ingrid; Siegel, Achim; Graf, Erika; Farin, Erik; Ihle, Peter; Köster, Ingrid; Stelzer, Dominikus; Mehl, Claudia; Schmitz, Jutta; Dröge, Patrik; Guenster, Christian; Klöss, Andreas; Vach, Werner; Geraedts, Max

VERSION 1 – REVIEW

REVIEWER	Melinda Davis Oregon Health & Science University Portland, OR, United States
REVIEW RETURNED	28-Aug-2018

GENERAL COMMENTS	This is an interesting protocol designed to evaluate the impact of a novel integrated health model focused on addressing population-health and the full spectrum of morbidities experienced by individuals in a specific geographic area, rather than on individual diseases. The authors propose to use to controls - 10 regions with a similar population and health care (primary controls) and a random sample of federal state insures (secondary controls). Although the abstract highlights the use of claims data for this analyses, the protocol text describes use of qualitative data and reviews in order to specific and 'non specific' indicators associated with quality of care. It would be good to briefly mention these steps in the abstract as well (page 2). Module A1 (page 7). Please provide additional details about how the potential indicators will be extracted. What is your analytic approach? Module A2 (page 7-8). What is the time-frame for your article review? How will you determine if an article will be included in your data set? What procedure will you use to abstract data from these articles, the guidelines, and the quality indicators databases? It appears you will organize by a JCAHO scheme, but how will you ensure the data is accurate and complete? Will you use multiple reviewers? Will you do a quality check for a sub-set of the articles or resources? Module A4 (page 8). What will happen if the participants disagree about an indicator that prior work identifies as critical for these types of analyses? Will you do a sensitivity analyses or "check" in some other way?

	Page 9 work package B. You mention using ICD-10 codes, but prior to 2015 ICD-9 codes were used. Do you also plan to do a cross-walk with ICD-9 codes given your data is from 2005-2015? Page 9. Please specify why 2006-2010 was identified as the start-up phase and 2011 to 2015 as the consolidation phase. Is this based on data, funding, or some other feature of the program. Please justify why these two cut-points are the best for your analyses (versus a 1-2 year start-up). Page 9. You describe the approach to continuous enrollment within a given year, but how will you handle changing enrollment over time? Do you think there will be an impact if someone has been enrolled in the intervention for a longer period of time? Will "years enrolled" be considered as a covariate? Will you use a propensity scoring/matching process to identify "similar" controls or control for potential variation in time enrolled in some way? Later on page 11 it appears you will be looking at cross-sections of patients. Please clarify how you will identify patients eligible for the analysis, is it just patients who were continuously enrolled in a specific year without consideration to prior coverage/time exposed to the intervention? Do you assume the trend analysis would control for this potential effect? Perhaps comment on this as a potential limitation.

REVIEWER	Wenxi Tang China Pharmaceutical University, China
REVIEW RETURNED	30-Aug-2018

GENERAL COMMENTS	How to evaluate integrated care reforms is still a question needs consensus. This study seems to be with considerable value, with the purpose to develop an evaluation system for integrated care reforms using claims data. In a general sense, I think the finding will benefit peer researchers in this area.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Melinda Davis
Institution and Country: Oregon Health & Science University, Portland, OR, United States Please state any competing interests or state ‘None declared’: None declared.

This is an interesting protocol designed to evaluate the impact of a novel integrated health model focused on addressing population-health and the full spectrum of morbidities experienced by individuals in a specific geographic area, rather than on individual diseases. The authors propose to use to controls - 10 regions with a similar population and health care (primary controls) and a random sample of federal state insures (secondary controls).

Q: Although the abstract highlights the use of claims data for this analyses, the protocol text describes use of qualitative data and reviews in order to specific and 'non specific' indicators associated with quality of care. It would be good to briefly mention these steps in the abstract as well (page 2).

A: We revised the respective sentence in the abstract as follows: Model-specific and ’non-specific’ indicators, adopted from the literature and enriched by focus group interviews will be used to evaluate the model’s effectiveness and potential unintended consequences by analysing health care utilisation in general.

Q: Module A1 (page 7). Please provide additional details about how the potential indicators will be extracted. What is your analytic approach?

Module A2 (page 7-8). What is the time-frame for your article review? How will you determine if an article will be included in your data set? What procedure will you use to abstract data from these articles, the guidelines, and the quality indicators databases? It appears you will organize by a JCAHO scheme, but how will you ensure the data is accurate and complete? Will you use multiple reviewers? Will you do a quality check for a sub-set of the articles or resources?

We added the following sentences to the manuscript:
Two independent reviewers will screen abstracts and full texts of articles, guidelines and QI databases for indicators suitable to measure the quality of programme-specific processes and outcomes of care. We will search the mentioned databases for English and German articles without time limit. Our focus will be on indicators assessing integrated care, health promotion and prevention. We will exclude indicators focusing on practice management, and in-hospital care. All potential indicators will be extracted and entered into a database using the scheme developed by the Joint Commission on Accreditation of Healthcare Organizations (1990) to describe the indicators. We will eliminate duplicates and check whether the respective indicator could be calculated using routine claims data of German sickness funds. The final list of suitable indicators will be assessed by the consensus panel (see A4).

Module A4 (page 8). What will happen if the participants disagree about an indicator that prior work identifies as critical for these types of analyses? Will you do a sensitivity analyses or "check" in some other way?

A: We agree with the reviewer that this situation can happen. We do not plan a sensitivity analysis, but trust in the expertise of our panel members not to remove the most relevant indicators.

Q: Page 9 work package B. You mention using ICD-10 codes, but prior to 2015 ICD-9 codes were used. Do you also plan to do a cross-walk with ICD-9 codes given your data is from 2005-2015?

A. In Germany, where the study takes place; ICD-10 code was introduced in the year 2000, therefore no cross coding from ICD-9 to ICD 10 is necessary.

We added the following information to the manuscript (page 9):
For the analysis, ICD-10 coded diagnoses – available since the year 2000- from out- and inpatient care are provided, further medical services according to the EBM Code (physician fee schedule), drug prescription with pharmaceutical registration number and linkage to Anatomical Therapeutic Chemical (ATC) classification and Defined Daily Dose (DDD), hospital stays with e.g. ICD-10 coded diagnoses, procedures ('OPS' Codes) and length of stay, benefits in kinds, information concerning inability to work (diagnosis, duration) and utilisation of long-term care.

Q: Page 9. Please specify why 2006-2010 was identified as the start-up phase and 2011 to 2015 as the consolidation phase. Is this based on data, funding, or some other feature of the program. Please justify why these two cut-points are the best for your analyses (versus a 1-2 year start-up).

A. The cut was driven by the contract of the program aiming at about 8.000 persons enrolled, which would offer the possibility for the management to make contracts with other sickness funds.

We added the following information to the manuscript (page 9):
The ICM GK project defined a number of 8.000 enrolled patients as a precondition to open the program for other sickness funds. This number was reached in 2011, therefore we will take the years 2006 to 2010 as the start-up phase, 2011 to 2015 as the consolidation phase. Furthermore, the

increase of enrollments has remarkably slowed down from 2011, showing that enrolment dynamics is another feature suggesting that we may differentiate those two development phases.

Page 9. You described the approach to continuous enrollment within a given year, but how will you handle changing enrollment over time? Do you think there will be an impact if someone has been enrolled in the intervention for a longer period of time? Will "years enrolled" be considered as a covariate? Will you use a propensity scoring/matching process to identify "similar" controls or control for potential variation in time enrolled in some way? Later on page 11 it appears you will be looking at cross-sections of patients. Please clarify how you will identify patients eligible for the analysis, is it just patients who were continuously enrolled in a specific year without consideration to prior coverage/time exposed to the intervention? Do you assume the trend analysis would control for this potential effect? Perhaps comment on this as a potential limitation.

A. As described in the introduction, the special feature of ICM –GK is the population-oriented approach. Aim of ICM-GK is to improve the health of a defined population, the quality of care for this whole population. Therefore, the evaluation also takes up this perspective and compares the ICM.-GKpopulation (irrespective of any enrollment) with populations in regions similar to the one ICM-GK is operating. This is described on page 9 (target and control population). All patients in the respective regions are eligible; whether they have to fulfill continuous enrollment with the sickness fund will be specified according to the particular indicators.

To better clarify this point we changed the sentence in the paragraph Target and control population as follows.
The target population consists of all AOK insurees living in the intervention region irrespective of any enrollment- – this results from the conception of the ICM-GK as a regional population-based health care system covering virtually all health care sectors and health conditions

However, it remains the issue that subjects immigrating into the intervention area will be less exposed to the intervention over time than those living there from the beginning. At the moment, we do not plan to take this explicitly into account in our analyses due to the following three reasons: First, diagnoses of patients will be only available on a quarterly basis, and hence we gave up very early the idea to base the analysis on following individual trajectories. Second, we expect that many indicators refer to annual rates requiring to include only patients who are present in the intervention region for the whole year or the majority of the year. Consequently, immigrating subjects will contribute only the next year after their immigration. Third, the initiatives undertaken as part of the intervention intend to influence the whole system, and not just the single patient. So immigrating patients may benefit from the intervention just from the first day in contrast to those living in the intervention region the whole time, who (actively) participate in changing the system. Consequently, simple approaches like including the exposure time as a covariate would not be adequate. Of course, we are aware of that using populations based on presence for a whole year in the intervention region leads to a selection, which may bias the results. For this reason we are currently performing preliminary analyses (without having access to any outcome data) looking only at the migration patterns of patients to get an idea about the magnitude of migration and the comparability of regions with respect to migration patterns. This will inform the final technical definition of the patient population to be fixed in the statistical analysis plan.

Reviewer: 2
Reviewer Name: Wenxi Tang
Institution and Country: China Pharmaceutical University, China Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below
How to evaluate integrated care reforms is still a question needs consensus. This study seems to be with considerable value, with the purpose to develop an evaluation system for integrated care reforms using claims data. In a general sense, I think the finding will benefit peer researchers in this area.

A *No comment necessary.*

VERSION 2 – REVIEW

REVIEWER	Wenxi Tang PhD Associate Professor China Pharmaceutical University, China
REVIEW RETURNED	03-Oct-2018

GENERAL COMMENTS	Dear authors, it is such a complicated analysis that you might have to test the effect of Gesundes Kinzigtal by using claims data, I would very much appreciated if you succeeded in carrying out this study. Though a lot of issues would exist no matter how specific you describe your working process and analysis strategy in this design, there are still ways to make them more clear before this study is conducted. First, I would appreciate if you could specify the methods you would adopt to test the differences among groups related to the designed types of indicators, and also the ways you wanted to use to control for the confounders that were obvious in this quasi-experiment. Second, you have mentioned the difficulties in obtaining desired indicators from the claims data only, then how you would expect to solve this problem in order to fully complete the evaluation based on the indicators generated by expert opinions? For the references, there are quite a few publications written in German which are hard to comprehend, and even from the years of publication, I can't say they are up-to-date. Finally, there are still places in your writing you might wanted to double check. For instance, in fig 1, the fokus group, I assumed it should focus group? A thorough re-reading with careful wording before publication is recommended here to avoid misunderstanding. Good luck with your study!

VERSION 2 – AUTHOR RESPONSE

Reviewer 2

Reviewer Name: Wenxi Tang PhD Associate Professor

Institution and Country: China Pharmaceutical University, China

Please state any competing interests or state ‘None declared’: None declared

Reviewer;

Dear authors, it is such a complicated analysis that you might have to test the effect of Gesundes Kinzigtal by using claims data, I would very much appreciated if you succeeded in carrying out this study. Though a lot of issues would exist no matter how specific you describe your working process and analysis strategy in this design, there are still ways to make them more clear before this study is conducted. First, I would appreciate if you could specify the methods you would adopt to test the differences among groups related to the designed types of indicators, and also the ways you wanted to use to control for the confounders that were obvious in this quasi-experiment.

Answer

We completely agree with the reviewer that it would be nice to report more about the statistical methods we intend to use. However, we do not use just one method, but - due to the complexity of the problem - a combination of different methods. We start with computing trends based on standardized prevalences for each year and region, and by this standardization we adjust for differences in age, gender, comorbidity, and social status. This is already mentioned in the manuscript. However, the details of this step are already non-trivial. Since we base the standardization on a logistic regression model and since we have rather huge sample sizes for some of the indicators, it is adequate to allow for non-linear relationships and interactions in dependence on the sample size. In addition, we need not only the estimates of the standardized prevalences for the further steps, but also their standard errors. Here, using the standard approach results in computing times growing in a quadratic manner with the sample size, hence requiring also a nonstandard approach. The trends themselves are then estimated using meta-analytic techniques from the estimated standardized prevalences. Finally, the population variation in trends among the control regions is again estimated by meta-analytic techniques in order to relate the difference between the intervention region and the mean of the control regions to the spread among the control regions.

Since this technical implementation is rather complex, we prefer to communicate only the basic conceptual approach here, and for this reason we would like to stick to the current, rough description in this protocol. We are currently preparing an additional paper describing the details of our approach, and this paper should supplement this study protocol.

1

Reviewer: Second, you have mentioned the difficulties in obtaining desired indicators from the claims data only, then how you would expect to solve this problem in order to fully complete the evaluation based on the indicators generated by expert opinions?

Answer

In the discussion section we mentioned that the use of routine data has some limitations (“Although investigating the effects of complex interventions by relying on routine data entails certain limitations, it remains a reasonable and acceptable procedure.”) This refers to the fact that each method has its specific limitations, but we argue with reference to comparable studies that there are good examples to utilize routine data for the evaluation of health care systems, especially when it is necessary to assess data over long time periods. We also pointed out that one characteristic of the indicators consented is the feasibility to assess the indicators with the information available in routine data. Due to this prerequisite, we feel confident to run this study based on routine data. These aspects are already explicated in the study protocol, therefore we do not see the need for further explications.

Reviewer: For the references, there are quite a few publications written in German which are hard to comprehend, and even from the years of publication, I can't say they are up-to-date.

Answer: We looked for comparable studies applying routine data. The dates of the publication range from 2008 to 2015. We now added a further study dealing with evaluation of a family doctorcentered care, published lately in English. It has been inserted as reference 44. (Karimova K, Uhlmann L, Hammer M, Guethlin C, Gerlach FM, Beyer M. The development of diabetes complications in GPcentered healthcare. Am J Manag Care. 2018 Jul;24(7):322-327.)

Reviewer: Finally, there are still places in your writing you might wanted to double check. For instance, in fig 1, the fokus group, I assumed it should focus group? A thorough re-reading with careful wording before publication is recommended here to avoid misunderstanding. Good luck with your study!

Answer

Thank you very much! We corrected fig 1 and hope that we identified all spelling mistakes.

VERSION 3 – REVIEW

REVIEWER	Wenxi Tang China Pharmaceutical University
REVIEW RETURNED	16-Oct-2018

GENERAL COMMENTS	Thanks for your response, and looking forward for this study to be carried out!