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.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Differences in inpatient performance of public general hospitals following implementation of a points-counting payment based on diagnosis-related groups: a robust multiple interrupted time series study in Wenzhou, China
AUTHORS	Zhu, Tingting; Chen, Chun; Zhang, Xinxin; Yang, Qingren; Hu, Yipao; Liu, Ruoyun; Zhang, Xiangyang; Dong, Yin

VERSION 1 – REVIEW

REVIEWER	Aronoff, Stephen Temple University, Pediatrics
REVIEW RETURNED	26-May-2023

GENERAL COMMENTS	I was asked to specifically review the statistical methods used in this study. The authors analyzed the data as an interrupted time series which seems appropriate given the design and outcomes. They employed a variation of traditional regression analysis, which is acceptable if the authors can demonstrate that the curves for both the pre and post groups for all of the measures included are both linear and independent. If the curves do not meet these conditions, the authors may wish to consider ARIMA analysis, assuming a suitable model can be fit to the data (Schaffer et al. BMC Medical Research Methodology (2021) 21:58 https://doi.org/10.1186/s12874-021-01235-8). Another approach that would require identical data from a comparable governmental division is difference in difference.

REVIEWER	Quantin, Catherine Centre Hospitalier Universitaire, Service de Biostatistique et Informatique Médicale
REVIEW RETURNED	14-Jun-2023

GENERAL COMMENTS	Comments to Authors: Thank you for giving me the opportunity to review this very interesting article studying the impact of the introduction of points counting payment on DRG on hospital performance indicators. There is not a large body of literature on the evaluation of this type of financing, which could make this work a real scientific advance. This study raises questions about changes in practices and the impact of the introduction of this type of financing reform in terms of the service provided to patients. The paper is globally well written. However, some major issues remain to be addressed, particularly in terms of methodology. The authors should stress the difficulty of interpreting the results, given that the start of the intervention was almost concomitant with the start of the Covid pandemic. Unless the authors redo the analyses using a

	difference-in-differences method, these results must be interpreted with great caution and cannot be used as a basis for conclusions. Major points: Introduction: -l. 97: The methods for financing which were used prior to the implementation of DRGs should be specified. -l. 144: specify what are primary, secondary and tertiary hospitals more clearly (briefly here and more precisely in the Material and Methods section). -l. 156: How is Wenzhou representative of the other cities involved in the experiment? What about the case-mix of this area? -l. 157: The objective should be better specified, respecting the principles of time-place-person (e.g. evaluation of hospital performance before and after the implementation of the reform between 2019 and 2021 in the Wenzhou region for all hospital stays). Material and methods: - Do you have any results for other hospitals that were not included in the intervention? In my opinion, only a comparison of the results obtained in hospitals benefiting from the payment policy intervention with other hospitals not included in this experiment could help interpret the results and their evolution over time. - Is it possible to control for a Covid effect? During the period when the experiment began, hospital activity was completely disturbed. We therefore do not know whether the changes measured are linked to changes attributable to the pandemic. - I strongly recommend the use of a difference-in-differences method for a sensitivity analysis at least, otherwise we can't be sure that the evolution over time is related to the intervention. Limitations - The authors should address the issues of quality of the data used. -The fact that the authors could not take account the COVID effect should be pointed out because it is a big issue. - The authors should highlight the current limitations of using DRGs. They should also write one or two remarks on the relevance of the evaluation indicators used (based in particular on DRGs). This would help to put the results of their evaluation into perspective. - The authors should also point out the limitations of an evaluation carried out very close to the time of the intervention, which might not produce the same results as other evaluations carried out at a later date. - The authors should emphasize the difficulty of interpreting the results, given that the start of the intervention was almost concomitant with the start of the Covid pandemic. Unless the authors redo the analyses using a difference-in-differences method, these results must be interpreted with great caution and cannot be used as a basis for conclusions. Abstract: -A sentence about the method used should be provided: the mention of the use of a difference-in-differences method must be included. -The number of episodes taken into account in the evaluation should be specified at the beginning of the results.

	- The authors should emphasize in their conclusion the difficulty of interpreting the results, given that the start of the intervention was almost concomitant with the start of the Covid pandemic. Minor points: Introduction: -l. 96-117: Perhaps reference should be made to the currently identified limits of the financing allowed by DRGs and what their implementation implies (coding time for example). -l. 119-120: Please include a reference. -l. 150: Please add a reference. Material and methods: -l. 172-184: I am not sure that this paragraph is essential to the understanding of the paper. A clearer explanation of the secondary and tertiary hospitals mentioned is required (both in the introduction and in points 183 and 184). -l. 234-244: this point seems very interesting but could be included in the strengths of the study. -l. 245-252: could be included as an appendix. Results: -The authors should keep the same number of digits after the comma. -l. 255: The authors should specify that the results are in table 1 at the beginning of the paragraph. -In my opinion, the descriptive and R-MITS results are too detailed, making them difficult to read. Perhaps the body of the text should simply highlight the main results and refer to the 2 tables, which are very clear. Discussion: - The authors should start with 2 sentences for the general effect observed, which seems positive from various points of view, before going into detail for each indicator. - In my opinion, between l. 334-l. 434 the text should focus more on the elements of the discussion (which are very interesting) than on the details of each indicator, which would make it possible to reduce the size of the text a little. Abstract: Design: “conducted to identify the different performance before and after intervention implementation”: I'm not quite sure what that means. The authors should rephrase this sentence. Conclusions: The results should be put into better perspective. Does this reform aim to reduce ALOS and TEI or not? The last sentence states that the reform does not reduce ALOS for tertiary institutions but it does for secondary institutions - this should be clarified.

VERSION 1 – AUTHOR RESPONSE

Reviewers' Comments:

Reviewer: 1
Dr. Stephen Aronoff, Temple University
Comments to the Author:

I was asked to specifically review the statistical methods used in this study. The authors analyzed the data as an interrupted time series which seems appropriate given the design and outcomes. They employed a variation of traditional regression analysis, which is acceptable if the authors can demonstrate that the curves for both the pre and post groups for all of the measures included are both linear and independent. If the curves do not meet these conditions, the authors may wish to consider ARIMA analysis, assuming a suitable model can be fit to the data (Schaffer et al. BMC Medical Research Methodology (2021) 21:58 https://doi.org/10.1186/s12874-021-01235-8). Another approach that would require identical data from a comparable governmental division is difference in difference.

Reply: Thanks very much for your kind comments. To conduct a formal test for the difference in the correlation structure for preintervention and postintervention phases, the robust interrupted time series model has used the ARIMA process to model the stochastic component. Therefore, we did not conduct an additional ARIMA analysis.

Reference:

[35] Cruz M, Bender M, Ombao H. A robust interrupted time series model for analyzing complex health care intervention data. Statistics in medicine. Dec 20 2017;36(29):4660-4676. doi:10.1002/sim.7443.

[Page 4665-4666] 3.2.2 Stochastic properties pre– and post–change point

The relevant reports of the model attached in Appendix 1 through 5, which can show that the curves for both the pre and post groups for all of the measures included are both linear and independent. We have illustrated it before going into detail for the model results in the Main Document.

[Lines 277-280, Page 8]

The residuals of the estimated model and auto-correlation functions are as expected and are shown in Appendix 1 through 5, demonstrating that the curves for both the pre- and post-groups for the measures included are linear and independent.

Reviewer: 2
Dr. Catherine Quantin, Centre Hospitalier Universitaire

Comments to Authors:
Thank you for giving me the opportunity to review this very interesting article studying the impact of the introduction of points counting payment on DRG on hospital performance indicators. There is not a large body of literature on the evaluation of this type of financing, which could make this work a real scientific advance. This study raises questions about changes in practices and the impact of the introduction of this type of financing reform in terms of the service provided to patients. The paper is globally well written. However, some major issues remain to be addressed, particularly in terms of methodology. The authors should stress the difficulty of interpreting the results, given that the start of the intervention was almost concomitant with the start of the Covid pandemic. Unless the authors redo the analyses using a difference-in-differences method, these results must be interpreted with great caution and cannot be used as a basis for conclusions.

Reply: Thanks very much for your kind and detailed comments. We have stressed the difficulty of interpreting the results. Besides, despite failing to use a difference-in-differences method, this study conducted R-MITS analysis to enhances the relationship between the evolution of results and the intervention.

[Lines 375-395, Page 11]

The findings presented here should be generalized with caution. The impact of COVID-19 reached Zhejiang province (to where Wenzhou is located) at the end of January; fortunately, the setback was temporary due to the implementation of strict prevention and control measures, and recovery gradually occurred after February 2020. Nevertheless, it may have unavoidably resulted in decreasing patient visits and hospital revenues, while increasing operation costs and medical efficiency, which may hinder the improvement of these selected indicators. We could have minimized the disruption of the COVID-19 by removing data from the early years of the policy, but we did not do so because it would have been too subjective. The R-MITS model can identify inferences regarding the estimation of global change points across units, rather than removing some series as decided by the study team. In such an external environment, the implementation of the DRGs' payment would be timely and may make up for the negative impact of the epidemic, which could motivate cost-control consciousness and management behaviors. Because under the motivation of DRGs' payment, hospitals have to make adjustments to strengthen service performance, otherwise the revenue loss would be huge. Theoretically, the policy can guide high-grade-level hospitals to focus on the treatment of more complex and severe diseases and the breakthrough of medical service level, while returning routine and common diseases to low-grade-level hospitals for treatment to improve the CMI indicator. In addition, motivating hospitals to improve medical levels and efficiency can promote other indicators. This study showed that the DRG policy had positive effects on hospital performance, even during COVID-19.

[Lines 396-403, Page 11]

Regarding the analysis method, we were unable to apply the difference-in-differences, because the basic medical insurance designated medical institutions of Zhejiang province were all required to implement DRGs' payment, including Wenzhou. In addition, we could not obtain data from the public general hospitals outside Zhejiang province where the DRGs or other payment reforms (like Diagnosis-Intervention Packet payment) were not implemented, under the process of the nationwide payment reform. Fortunately, R-MITS is an optimized version of ITS, whose strengths can enhance the relationship between the evolution of results and the DRGs' policy intervention.

Major points:

Introduction:

1. Line 97: The methods for financing which were used prior to the implementation of DRGs should be specified.

Reply: Thank you for the detailed review. We have illustrated the payment method for financing which was used prior to the implementation of DRGs.

[Lines 57-59, Page 2]

Before DRGs' payment, the inpatient medical services in China mainly adopted Fee-For-Service, which easily causes the over-treatment and restrains the improvement of service performance.

2. Line 144: specify what are primary, secondary and tertiary hospitals more clearly (briefly here and more precisely in the Material and Methods section).

Reply: Thanks very much for your kind comments. We have revised and added the relative text.

[Lines 116-119, Page 3]

Chinese Health Commission divides medical institutions into three levels according to their tasks and functions to form a hierarchical medical system: primary, secondary, and tertiary hospitals.

[Lines 175-181, Page 5]

According to Chinese hospital classification and management standards, primary hospitals are hospitals that provide preventive, medical, healthcare, and rehabilitation services directly to a community of a certain population. Secondary hospitals are regional hospitals that provide comprehensive medical and health services to multiple communities and undertake certain teaching and research. Tertiary hospitals are hospitals above the regional level that provide high-level specialized medical services to several regions and perform higher medical education and research.

3. Line 156: How is Wenzhou representative of the other cities involved in the experiment? What about the case-mix of this area?

Reply: Thanks for your kind suggestions. We have interpreted the reason that the conditions of Wenzhou as a representative involved in the experiment and introduced the case-mix of this area.

[Lines 154-159, Page 4]

Wenzhou City lies in the southeast of Zhejiang province. In 2021, the resident population was 9.6450 million—the second most populous city in Zhejiang. It has a wealth of successful experience in reform and strong comprehensive medical service, a similar DRGs' policy implementation process and healthcare system structure to other cites, which act as a representative for estimating the effects of the PCP on public general hospitals.

[Lines 161-168, Page 5]

In 2019, the Wenzhou Healthcare Security Administration conducted disease grouping and cost measurement for the past three years. The base data was collected from 161 medical institutions in the city that offer hospitalization. After grouping by DRGs, hospitalization cases were divided into 912 DRGs. Among them were 55 DRGs where the number of cases was ≤5; 857 DRGs where number of cases was＞5 and their coefficient of variation was <1; reduction in variance was 75.04% and the cost differences across groups were all >20%, which reached the national technical specifications for DRGs grouping and payment.

4. Line 157: The objective should be better specified, respecting the principles of time-place-person (e.g., evaluation of hospital performance before and after the implementation of the reform between 2018 and 2021 in the Wenzhou region for all hospital stays).

Reply: Thanks for your kind suggestion. The objective has been better specified under your requirement.

[Lines 149-151, Page 4]

This study takes Wenzhou in Zhejiang province as a sample city to evaluate hospital performance before and after the implementation of the DRGs' reform between 2018 and 2021 in public general hospitals from the perspective of institution levels.

Material and methods:

5. Do you have any results for other hospitals that were not included in the intervention? In my opinion, only a comparison of the results obtained in hospitals benefiting from the payment policy

intervention with other hospitals not included in this experiment could help interpret the results and their evolution over time.

Reply: Thank you for the detailed review. We have not any results for other hospitals that were not included in the intervention. Zhejiang' basic medical insurance designated medical institutions were all required to implement the DRGs' payment, including the hospitals in the sample region, thus the intervention is all implemented in these hospitals. In addition, we could not obtain data from the public general hospitals outside Zhejiang province where the DRGs or other payment reforms (like Diagnosis-Intervention Packet payment) were not implemented, under the process of the nationwide payment reform.

6. Is it possible to control for a Covid effect? During the period when the experiment began, hospital activity was completely disturbed. We therefore do not know whether the changes measured are linked to changes attributable to the pandemic.

Reply: Thanks for your great suggestion on improving the accessibility of our manuscript. We have interpreted the Covid effect, which is fully described in the penultimate paragraph of the Discussion.

[Lines 375-395, Page 11]

The impact of COVID-19 reached Zhejiang province (to where Wenzhou is located) at the end of January; fortunately, the setback was temporary due to the implementation of strict prevention and control measures, and recovery gradually occurred after February 2020. Nevertheless, it may have unavoidably resulted in decreasing patient visits and hospital revenues, while increasing operation costs and medical efficiency, which may hinder the improvement of these selected indicators. We could have minimized the disruption of the COVID-19 by removing data from the early years of the policy, but we did not do so because it would have been too subjective. The R-MITS model can identify inferences regarding the estimation of global change points across units, rather than removing some series as decided by the study team. In such an external environment, the implementation of the DRGs' payment would be timely and may make up for the negative impact of the epidemic, which could motivate cost-control consciousness and management behaviors. Because under the motivation of DRGs' payment, hospitals have to make adjustments to strengthen service performance, otherwise the revenue loss would be huge. Theoretically, the policy can guide high-grade-level hospitals to focus on the treatment of more complex and severe diseases and the breakthrough of medical service level, while returning routine and common diseases to low-grade-level hospitals for treatment to improve the CMI indicator. In addition, motivating hospitals to improve medical levels and efficiency can promote other indicators. This study showed that the DRG policy had positive effects on hospital performance, even during COVID-19.

7．I strongly recommend the use of a difference-in-differences method for a sensitivity analysis at least, otherwise we can't be sure that the evolution over time is related to the intervention.

Reply: Thanks for your kind recommendation. We were also keen to use a difference-in-differences method for sensitivity analysis, but we were unable to design the control group that was not affected by the same intervention. The interprets are added in the last paragraph of the Discussion.

[Lines 396-403, Page 11]

Regarding the analysis method, we were unable to apply the difference-in-differences, because the basic medical insurance designated medical institutions of Zhejiang province were all required to implement DRGs' payment, including Wenzhou. In addition, we could not obtain data from the public general hospitals outside Zhejiang province where the DRGs or other payment reforms (like Diagnosis-Intervention Packet payment) were not implemented, under the process of the nationwide payment reform. Fortunately, R-MITS is an optimized version of ITS, whose strengths can enhance the relationship between the evolution of results and the DRGs' policy intervention.

Limitations:

8．The authors should address the issues of quality of the data used.

Reply: Thank you for the detailed review. We have illustrated the issues of quality of the data used both in Material and methods, and the last paragraph of Discussion.

[Lines 189-191, Page 5]

The data were collected from the Zhejiang Hospital Quality Management and Performance Evaluation Platform, which was developed by the Zhejiang Provincial Health Commission. The platform launched in 2017 and, reviewed the quality of the data uploaded by each hospital by logic verification and key quality control index monthly before evaluating the performance.

[Lines 406-409, Page 12]

The DRGs' indicators can homogeneously compare the medical performance, but the main drawback is that they are more dependent on the quality of the data than traditional indicators, especially the quality of the homepage of inpatient medical records, including coding quality.

9. The fact that the authors could not take account the COVID effect should be pointed out because it is a big issue.

Reply: Thanks for your great suggestion on improving the accessibility of our manuscript. Despite the inevitable impact of COVID-19 on hospitals, the design of this study succeeded in revealing some anticipated and positive effects of the DRGs' policy on hospitalization performance, making the study meaningful.

10. The authors should highlight the current limitations of using DRGs. They should also write one or two remarks on the relevance of the evaluation indicators used (based in particular on DRGs). This would help to put the results of their evaluation into perspective.

Reply: Thanks for your great suggestion. We have highlighted the current limitations of using DRGs and written remarks on the relevance of the evaluation indicators used.

[Lines 406-413, Page 12]

DRGs' indicators can homogeneously compare the medical performance, but their calculations are more dependent on the quality of the homepage of inpatient medical records than traditional indicators, especially disease coding quality. Thus, previous studies that used the DRGs' indicators were almost conducted in above-primary-level hospitals, include this study. CMI reflects the average level of medical technology and complexity of the diseases attended to by each hospital; CEI or TEI reflect the relative efficiency of CPH or ALOS, which are calculated based on DRGs.

11. The authors should also point out the limitations of an evaluation carried out very close to the time of the intervention, which might not produce the same results as other evaluations carried out at a later date.

Reply: Thanks for your kind suggestion. We have added it in the last paragraph of the Discussion.

[Lines 413-415, Page 12]

This evaluation was conducted close to the intervention's time, which might not produce the same results as other evaluations carried out at a later date.

12. The authors should emphasize the difficulty of interpreting the results, given that the start of the intervention was almost concomitant with the start of the Covid pandemic. Unless the authors redo the analyses using a difference-in-differences method, these results must be interpreted with great caution and cannot be used as a basis for conclusions.

Reply: Thanks for your valuable suggestion. We have emphasized the difficulty of interpreting the results under the Covid pandemic and why not using a difference-in-differences method.

[Lines 375-395, Page 11]

The findings presented here should be generalized with caution. The impact of COVID-19 reached Zhejiang province (to where Wenzhou is located) at the end of January; fortunately, the setback was temporary due to the implementation of strict prevention and control measures, and recovery gradually occurred after February 2020. Nevertheless, it may have unavoidably resulted in decreasing patient visits and hospital revenues, while increasing operation costs and medical efficiency, which may hinder the improvement of these selected indicators. We could have minimized the disruption of the COVID-19 by removing data from the early years of the policy, but we did not do so because it would have been too subjective. The R-MITS model can identify inferences regarding the estimation of global change points across units, rather than removing some series as decided by the study team. In such an external environment, the implementation of the DRGs' payment would be timely and may make up for the negative impact of the epidemic, which could motivate cost-control consciousness and management behaviors. Because under the motivation of DRGs' payment, hospitals have to make adjustments to strengthen service performance, otherwise the revenue loss would be huge. Theoretically, the policy can guide high-grade-level hospitals to focus on the treatment of more complex and severe diseases and the breakthrough of medical service level, while returning routine and common diseases to low-grade-level hospitals for treatment to improve the CMI indicator. In addition, motivating hospitals to improve medical levels and efficiency can promote other indicators. This study showed that the DRG policy had positive effects on hospital performance, even during COVID-19.

[Lines 396-403, Page 11]

Regarding the analysis method, we were unable to apply the difference-in-differences, because the basic medical insurance designated medical institutions of Zhejiang province were all required to implement DRGs' payment, including Wenzhou. In addition, we could not obtain data from the public general hospitals outside Zhejiang province where the DRGs or other payment reforms (like Diagnosis-Intervention Packet payment) were not implemented, under the process of the nationwide

payment reform. Fortunately, R-MITS is an optimized version of ITS, whose strengths can enhance the relationship between the evolution of results and the DRGs' policy intervention.

Abstract:

13. A sentence about the method used should be provided: the mention of the use of a difference-in-differences method must be included.

Reply: Thanks for your kind suggestion. We were also keen to use a difference-in-differences method for sensitivity analysis, but we were unable to design the control group that was not affected by the same intervention. The interprets are revised in last paragraph of the Discussion. We have illustrated the method of this study in the Design of Abstract: a robust multiple-interrupted time-series (R-MITS) method.

[Lines 11-12, Page 1]

Design: A longitudinal study using a robust multiple interrupted time-series model to evaluate service performance following policy implementation.

[Lines 396-403, Page 11]

Regarding the analysis method, we were unable to apply the difference-in-differences, because the basic medical insurance designated medical institutions of Zhejiang province were all required to implement DRGs' payment, including Wenzhou. In addition, we could not obtain data from the public general hospitals outside Zhejiang province where the DRGs or other payment reforms (like Diagnosis-Intervention Packet payment) were not implemented, under the process of the nationwide payment reform. Fortunately, R-MITS is an optimized version of ITS, whose strengths can enhance the relationship between the evolution of results and the DRGs' policy intervention.

14. The number of episodes taken into account in the evaluation should be specified at the beginning of the results.

Reply: Thanks for your kind suggestion. We have specified the number of episodes taken into account in the evaluation at the beginning of the Results of Abstract.

[Lines 19-21, Page 1]

The impact of COVID-19 was temporary, which reached Zhejiang province at the end of January 2020, and had been contained quickly, benefiting from the Chinese strict control measures.

15. The authors should emphasize in their conclusion the difficulty of interpreting the results, given that the start of the intervention was almost concomitant with the start of the Covid pandemic.

Reply: Thanks for your suggestion. We have emphasized the difficulty of interpreting the results.

[Lines 29-30, Page 1]

This study showed a positive effect of the DRGs' policy in Wenzhou, even during COVID-19.

******

Minor points:

Introduction:

1. Lines 96-117: Perhaps reference should be made to the currently identified limits of the financing allowed by DRGs and what their implementation implies (coding time for example).

Reply: Thanks very much for your kind suggestion. We have supplied the currently identified limits of the financing allowed by DRGs and what their implementation implies.

[Lines 84-89, Page 3]

The DRGs' payment implementation requires hospitals to be proactive regarding the quality of medical records (like more long coding time to ensure the quality of case grouping) and an improvement in the performance of their healthcare services, to avoid losses after the policy implementation. Under the given pricing standard, hospitals are responsible for their losses in cases where medical costs exceed the pricing standard.

2. Lines 119-120: Please include a reference.

Reply: Thanks very much for your valuable suggestion. We have checked the literature carefully and added two references on “previous studies may have neglected to include the time until the policy’ effects manifested” into the Introduction part in the revised manuscript.

[Lines 113-114, Page 3]

[22] Jiang Q, Xu Z, Yu L, Zhou H, Zhang Z. Impact Analysis on the Expenses of C-DRG Pricing and Payment Reform on Medical Facilities in Sanming City. Chinese Health Economics. 2021;40(04):13-16.

[23] Fang J, Liu L, Peng Y, Tao H. Analysis of the Impact of CHS-DRG Payment Reform on Hospital Operation CHS-DRG. Health Economics Research. 2022;39(05):67-71.

3. Line 150: Please add a reference.

Reply: Thanks for the kind suggestion. We have checked the literature carefully and added a reference on “Previous studies have analyzed them as subgroups in healthcare” into the Introduction part in the revised manuscript.

[Lines 120-121, Page 3]

[25] Gao L, Shi L, Meng Q, Kong X, Guo M, Lu F. Effect of healthcare system reforms on public hospitals' revenue structures: Evidence from Beijing, China. Soc Sci Med. Aug 2021; 283:114210.

Material and methods:

4. Lines 172-184: I am not sure that this paragraph is essential to the understanding of the paper. A clearer explanation of the secondary and tertiary hospitals mentioned is required (both in the introduction and in points 183 and 184).

Reply: Thanks very much for your kind suggestion. We have deleted this paragraph Lines 172-184 below.

“First, in this region, the hospitals established DRGs’ work teams, which could negotiate with the Healthcare Security Administration to safeguard the hospitals’ interests. Such teams were generally composed of the dean, who oversaw the DRG’ performance, the medical insurance department, and medical record quality-control department managers. Second, interior DRGs’ data-monitoring

information systems were introduced to strengthen hospital-wide disease groups monitoring, including monitoring of profit and loss and key indicators. Third, quality control of medical records was strengthened, especially the quality of code uploads, which was one of the prerequisites for obtaining medical insurance funds through DRGs' payment. Fourth, many hospitals launched training for clinicians across departments and specialized guidelines, which were also vital for the reform of the healthcare payment system, and helped to standardize diagnoses."

Besides, we have added a clearer explanation of the secondary and tertiary hospitals mentioned in the Introduction and before the previous Lines 183-184 "Above-primary-level public general hospitals in Wenzhou amounted to 22 institutions, which consisted of 8 tertiary hospitals and 14 secondary hospitals.".

[Lines 116-119, Page 3]

Chinese Health Commission divides medical institutions into three levels according to their tasks and functions to form a hierarchical medical system: primary, secondary, and tertiary hospitals.

[Lines 175-181, Page 5]

According to Chinese hospital classification and management standards, primary hospitals are hospitals that provide preventive, medical, healthcare, and rehabilitation services directly to a community of a certain population. Secondary hospitals are regional hospitals that provide comprehensive medical and health services to multiple communities and undertake certain teaching and research. Tertiary hospitals are hospitals above the regional level that provide high-level specialized medical services to several regions and perform higher medical education and research.

16. Lines 234-244: this point seems very interesting but could be included in the strengths of the study.

Reply: Thanks for your suggestion. We have added these points to the Strengths of the study.

[Lines 39-46, Page 2]

A robust multiple interrupted time series (R-MITS) model does not assume that the effect of intervention is instantaneous and allows the presence of a potentially lagged (or anticipatory) effect, which match the characteristics of the DRGs' policy.

R-MITS estimates the global change-point, at which the effect of the intervention initiates for the entire health system, rather than removing some points as decided by the study team.

R-MITS can capture mean differences in variability and auto-correlation of time-series changes, improve validity and reduces bias to obtain more accurate estimates.

17. Lines 245-252: could be included as an appendix.

Reply: Thanks for your kind suggestion. We have put the below corresponding contents and added the other details of model added into Appendix A.

[Appendix A in Supplemental Material]

Let y_{it} denote the outcome of interest for unit i at time t ,with $i \in \{1,...,N\}$, $t \in \{1,...,n_i\}$, and n_i denoting the time series length for unit i. Then the general regression is defined as

$$y_{it} = \mu_{it} + \varepsilon_{it}$$

where μ_{it} is the mean function and ε_{it} is the stochastic process that model's fluctuations around the mean functions and auto-correlation within the time series. The mean function of outcome is

$$\mu_{it} = \begin{cases} \beta_{i0}^{\tau} + \beta_{i1}^{\tau} t, & t < \tau \\ (\beta_{i0}^{\tau} + \delta_i^{\tau}) + (\beta_{i1}^{\tau} + \Delta_i^{\tau}) t, & t \geq \tau \end{cases}$$

where β_{i0}^{τ} denotes the intercept of the mean function prior to the change-point, β_{i1}^{τ} denotes the slope of the outcome prior to the change-point, $\beta_{i0}^{\tau} + \delta_i^{\tau}$ is the intercept of the post-intervention phase, $\beta_{i1}^{\tau} + \Delta_i^{\tau}$ is the slope of the post-intervention phase for the outcome in unit i, and τ denotes the global over-all-unit change-point of the response. In the case with only one unit, τ denotes the change-point for that one-time series. If $\delta_i^{\tau} + \Delta_i^{\tau} = 0$, then there is no change in the mean function of unit i before and after τ. Further model details about the estimation procedure of R-MITS are described in Cruz et al (2021).

Results:

18. The authors should keep the same number of digits after the comma.

Reply: Thanks very much for your detailed comments. We have kept the same number of digits after the comma of the study results, and here we did not list the changes but marked them in the revised paper.

19. Line 255: The authors should specify that the results are in Table 1 at the beginning of the paragraph.

Reply: Thanks for your suggestion. We have specified that the results are in Table 1 at the beginning of the paragraph.

[Lines 252-253, Page 7]

The yearly outcomes for every target indicator for the period of 2018–2021 (representing pre- and post-implementation) are shown in Table 1.

20. In my opinion, the descriptive and R-MITS results are too detailed, making them difficult to read. Perhaps the body of the text should simply highlight the main results and refer to the 2 tables, which are very clear.

Reply: Thanks very much for your kind comments. We have highlighted the main results, these changes will not influence the content of the paper, and here we did not list the changes but marked in the revised paper.

Discussion:

21. The authors should start with 2 sentences for the general effect observed, which seems positive from various points of view, before going into detail for each indicator.

Reply: Thanks for your great suggestion. We have added 2 sentences for the general effect observed before going into detail for each indicator in the Discussion.

[Lines 306-309, Page 9]

With limited and stretched healthcare resources, it is important for hospitals to further improve the performance of their medical services after the new policy implementation. In general, the results from all samples and subgroups seem positive from various perspectives.

22. In my opinion, between l. 334-l. 434 the text should focus more on the elements of the discussion (which are very interesting) than on the details of each indicator, which would make it possible to reduce the size of the text a little.

Reply: Thanks very much for your kind comments. We have focused more on the elements of the discussion (which are very interesting) than on the details of each indicator. These changes will not influence the content and framework of the paper, and here we did not list the changes but marked in the revised paper.

Abstract:

23. Design: “conducted to identify the different performance before and after intervention

implementation”: I'm not quite sure what that means. The authors should rephrase this sentence.

Reply: Thanks for your detailed suggestion. We have rephrased this sentence.

[Lines 11-12, Page 1]

Design: A longitudinal study using a robust multiple interrupted time-series model to evaluate service performance following policy implementation.

24. Conclusions: The results should be put into better perspective. Does this reform aim to reduce ALOS and TEI or not? The last sentence states that the reform does not reduce ALOS for tertiary institutions but it does for secondary institutions - this should be clarified.

Reply: Thanks for your great suggestion on improving the accessibility of our manuscript. We have elaborated again the Conclusions of the Abstract not focusing on the indicators, which become more better perspective.

[Lines 29-34, Page 1]

This study showed a positive effect of the DRGs’ policy in Wenzhou, even during COVID-19. It can motivate public general hospitals to improve the comprehensive capacity and mitigate discrepancies the efficiency in the expense of treating similar diseases. Policymakers are interested in whether the reform successfully stimulates hospitals to strengthen the internal impetus to improve performance, which is supported by this study.

VERSION 2 – REVIEW

REVIEWER	Quantin, Catherine Centre Hospitalier Universitaire, Service de Biostatistique et Informatique Médicale
REVIEW RETURNED	06-Nov-2023

GENERAL COMMENTS	The authors have replied to all my comments.