nature portfolio

Peer Review File

Regorafenib plus sintilimab as a salvage treatment for microsatellite stable metastatic colorectal cancer: a single-arm, open-label, phase II clinical trial

REVIEWER COMMENTS

Reviewer #1 (CRC, clinical trial):

This trial addresses an important unmet need in pMMR/MSS colorectal cancer.

The Authors should be congratulated on this study>

Though, I have a number of comments that the Authors should consider addressing to improve the quality of the study.

Major comments

The main weakness of this study is the absence of a randomised design with a control arm including regorafenib alone. While the results are encouraging, the conclusions about the inter-trial comparison with the randomised phase III CORRECT trial should be cautiously interpreted. In this regard, it is surprising that the Authors do not acknowledge any limitation (including the single-arm design, the risk of patient selection bias, the inclusion of a Chinese-only population, etc.) in the manuscript While data from initial single-arm trials of multikinase inhibitors with immune checkpoint inhibitors in Eastern patients were especially promising, these were confirmed neither in many single-arm studies in Werstern countries nor in randomised trials (ie, LEAP-017). The manuscript would require extensive linguistic review.

Minor comments

- “Systemic treatment for mCRC typically involves chemotherapy with oxaliplatin, irinotecan, or 5-FU-based doublet or triplet regimens”. This sentence is misleading (oxaliplatin and irinotecan are not generally used as monotherapy) and should be amended. Similarly the sentence “Targeted therapy may also be added to this paradigm” should be amended as well (addition of targeted therapy to cytotoxic chemotherapy is actually recommended in international guidelines).

- Neither regorafenib nor trifluridine/tipiracil can be described as “novel agents”.

- Regorafenib have more targets than those listed in the introduction section. In this regard, referring to the study treatment the authors talk about anti-angiogenic therapies plus immune checkpoint inhibitors throughout the manuscript, but the synergistic effect between immunotherapy and reforafenib may not necessarily (or exclusively) be secondary to the angiogenic activity of this ,ultikinase inhibitor.

- “Prognostic phase II study” does not mean anything.

- “The minimum required sample size was calculated to be 68 patients. Considering a dropout rate of 20%, the final sample size was determined to be 82 patients. Consequently, a total of 103 patients were enrolled in the study”. Based on the sample size calculation as described in this sentence, it is not clear how the authors end up having to recruit 103 patients.

- Good to clarify the technique used for the assessment of microsatellite stability (IHC, PCR, both?)

- The title wording of section “2.3 Description of the results” should be amended as this section does not report results, but it describes the evaluation criteria for the study endpoints.

- History of previous surgery was an independent prognostic factor. Does this refer to the surgery of the primary tumour or metastases?

- There seems to be a substantial difference between mPFS (4.1 months) and mOS (14.1 months). This may suggest an impact of post-study treatments or a patient selection bias. The inclusion of a prognostically favourable population is suggested by the fact that none of the independent prognostic factors for PFS is independently prognostic for OS.

- To clarify if the 20 tumour samples from the primary tumours were taken before study entry or at the time of diagnostic biopsy or initial resection.

- Not sure I understand the following sentence: “A comparison of tumorous paracancerous tissues revealed that the expression patterns of the upregulated genes in the treatment-sensitive group (DYRK2, ID4) were under expressed in the tumor tissues, and the genes that were downregulated in the treatment-sensitive group (TRMT112, ULBP1) were highly expressed in the tumor tissues (Figure 7-C)”.

- Reference 8 appears inappropriate as it includes a study in hepatocellular carcinoma rather than colorectal cancer.

- Recent and relevant studies should be discussed in the introduction and/or discussion section, such as the ATEZOTribe trial. Similarly to what the authors did, this study also evaluated the Immunoscore IC, taking into account not only the density of CD3/CD8 cells but also their spatial distribution and proximity to the tumour cells.

- Table 1 is reported twice with slighlty different headings. Please clarify.

- In Figure 1, 8 patients are indicated as being excluded but reasons for exclusions are reported only for 5 of them.

- Numbers at risk should be included in all figures

- Forest plots should report the 95%CI bars for all variables (including those which are used as reference).

Reviewer #2 (cancer therapy, clinical trial):

Liu R et al. presents a phase II study evaluating the efficacy and safety of the combination of regorafenib and sintilimab in patients with MSS mCRC. This study suggests promising efficacy with median OS of 14.1 months and 21.4% of ORR in patients who show typically poor response to immunotherapy.

1. The study reported a sample size of 82 patients but enrolled 103. Please clarify the reason for this discrepancy and assess the potential impact on the study's power.

2. The data seems to be that RAS/RAF status might be a prognostic rather than predictive factor. Moreover, the findings contradict those of the LEAP-017 trial. Please elaborate on this discrepancy and discuss its implications.

3. Please specify whether the reported ORR is confirmed or unconfirmed.

4. Figure 5A shows a case classified as PR with less than 30% tumor shrinkage.

5. Please analyze and report efficacy according to PD-L1 expression levels (e.g., CPS or TPS).

6. Please provide detailed information on serious adverse events (SAEs) to comprehensively assess the safety profile of the combination therapy.

7. The authors claim that Chinese patients may benefit more from the combination therapy. Please provide evidence to support this assertion and discuss potential underlying mechanisms.

8. Given the relevance of the LEAP-017 trial, a detailed comparison of the results and implications is warranted.

Minor

1. Cite the CONCUR trial as a reference for the assessment criteria threshold.

2. Ensure consistent use of decimal points throughout the manuscript.

3. In conclusion, revise "particularly in patients with RAS wild-type gene" to "particularly in patients with RAS/RAF wild-type gene."

Reviewer #3 (Biostatistician, clinical trial):

In this phase II trial the authors report the efficacy and safety results from regorafenib plus sintilimab as salvage treatment for microsatellite stable (MSS) in mCRC patients. There are several

a few flaws with this trial, mainly the choice of the primary endpoint and the incorrect use of the test statistic to determine the sample size. It would be helpful to see the protocol.

1) Primary Endpoint of OS in a Single-Arm Phase II Trial: Selecting overall survival (OS) as the primary endpoint in a single-arm Phase II trial is unconventional. Typically, OS is reserved for randomized Phase III trials, as comparing survival outcomes requires a control arm. The authors need to clearly justify the choice of OS and explain how they plan to interpret this endpoint in the absence of a comparator group.

2) It is not clear how the sample was computed in this single arm phase II trial. The authors refer to the single log rank test. This reviewer is not familiar with this test as the log-rank test Can they elaborate? What was the alternative hypothesis?

3) In Table 1, can you replace ECOG Score with "ECOG PS" and also uncouple ECOG PS as 0, 1, to present these two groups separately? Also, where there are several possible answer, indicate that the categories are non-mutually exclusive (such as prior previous treatment)

4) For the ORR, were the PR and CR based on confirmed response?

5) How many deaths were observed? Was the 55 deaths due to treatment or disease progression?

6) For all the 95% CI, please report 1 digit after the decimal point.

7) In Table 2, can you present grade 3, 4, and 5 as individual categories?

8) The results from the multivariable analysis show wide 95% CI, which is expected as the number of deaths is small. The estimated hazard ratios are unstable and I would refrain from reporting these results.

9) For the Kaplan-Meier curves (2A, 3A-3B, 4), label the x axis properly. Is it time since enrollment (months?)

10) Remove the p-values from Figures 2B & 4B; as no prior hypotheses were specified and it would be hard to interpret the results.

11) What is the purpose of Figure 5A?

12) The resolutions of some of the figures are weak (such as Figure 7).

13) Please check for typos: Trail should be trial, etc.

Reviewer #1 (CRC, clinical trial):

This trial addresses an important unmet need in pMMR/MSS colorectal cancer.
The Authors should be congratulated on this study>
Though, I have a number of comments that the Authors should consider addressing to improve the quality of the study.

Major comments
The main weakness of this study is the absence of a randomised design with a control arm including regorafenib alone. While the results are encouraging, the conclusions about the inter-trial comparison with the randomised phase III CORRECT trial should be cautiously interpreted. In this regard, it is surprising that the Authors do not acknowledge any limitation (including the single-arm design, the risk of patient selection bias, the inclusion of a Chinese-only population, etc.) in the manuscript. While data from initial single-arm trials of multikinase inhibitors with immune checkpoint inhibitors in Eastern patients were especially promising, these were confirmed neither in many single-arm studies in Werstern countries nor in randomised trials (ie, LEAP-017). The manuscript would require extensive linguistic review.
Response: When developing new treatment options, especially for patients with poor prognosis and limited treatment options, such as patients with MSS metastatic colorectal cancer, single-arm trial designs are often used to rapidly evaluate the efficacy and safety of treatment. The main reason for choosing the single-arm design is that it can obtain preliminary data more quickly than randomized controlled trials, especially for refractory diseases, and can provide patients with possible treatment options more quickly. Although this study did not establish a randomized control group, we explained the treatment effect by comparing it with historical control data. Compared with the third-line standard treatment for metastatic colorectal cancer (regorfenib monotherapy, TAS-102 monotherapy), the results of this study were improved regardless of ORR, median PFS, and median OS, which can be used as a reference. On the other hand, compared with previous trials of multikinase inhibitors with immune checkpoint inhibitors (REGONIVO, REGOMUNE, REGOTORI, etc.), the results were similar.
We used strict inclusion criteria to minimize selection bias. This includes a clear definition of patient characteristics (e.g. microsatellite stable state, previous treatment history, etc.) to ensure that the patients recruited are representative of the target population. In the process of recruiting participants, we re-evaluated the event rate and adjusted the sample size calculation based on actual data to ensure the statistical power of the study and the reliability of the results. We have supplemented the limitations of this study in the Discussion section of the manuscript. At the same time, we will further explain why the current study design was chosen and how this can be improved in future studies.
Although the result of the Phase III study LEAP-017 was negative, the subgroup

analysis suggested that the Asian population was more likely to benefit from this treatment mode. The patients in this study were all Chinese, which may be the population benefiting from this treatment mode, and the Phase III randomized controlled study needs to be further conducted in this population.

Minor comments

- “Systemic treatment for mCRC typically involves chemotherapy with oxaliplatin, irinotecan, or 5-FU-based doublet or triplet regimens”. This sentence is misleading (oxaliplatin and irinotecan are not generally used as monotherapy) and should be amended. Similarly the sentence “Targeted therapy may also be added to this paradigm” should be amended as well (addition of targeted therapy to cytotoxic chemotherapy is actually recommended in international guidelines).

Response: This has been revised in the manuscript.

-Neither regorafenib nor trifluridine/tipiracil can be described as “novel agents”.

Response: This has been revised in the manuscript.

-Regorafenib have more targets than those listed in the introduction section. In this regard, referring to the study treatment the authors talk about anti-angiogenic therapies plus immune checkpoint inhibitors throughout the manuscript, but the synergistic effect between immunotherapy and reforafenib may not necessarily (or exclusively) be secondary to the angiogenic activity of this ,ultikinase inhibitor.

Response: Previous basic studies have confirmed that abnormal tumor angiogenesis can lead to inhibitory immune microenvironment, so anti-angiogenesis therapy can reverse the inhibitory immune microenvironment and further increase the efficacy of immunotherapy. It is not clear whether other sites of regorafenib can increase immunotherapy efficacy. Your comments are so constructive and our basic research is also exploring this area. We have supplemented the content in the discussion section of the manuscript.

-“Prognostic phase II study” does not mean anything.

Response: We are sorry that it’s a clerical error and this has been revised in the manuscript.

-“The minimum required sample size was calculated to be 68 patients. Considering a dropout rate of 20%, the final sample size was determined to be 82 patients. Consequently, a total of 103 patients were enrolled in the study”. Based on the sample size calculation as described in this sentence, it is not clear how the authors end up having to recruit 103 patients.

Response: This study was an exploratory single-arm study, which was calculated using the one-sample log rank test. Based on previous literature and related research, it was assumed that the addition of sintilimab to regorafenib could

elevate the median OS from 8.8 months to 13.5 months, with an anticipated hazard ratio of 0.65 and event rate of 70%. This assumption was used to calculate the sample size to ensure that, with a significance level (α) of 0.5 and a power of 0.80, the number of events was 48, the initially calculated sample size was 68 participants. Considering a dropout rate of 20%, the initially sample size was determined to be 82 patients.

However, during the actual study, we observed that the event rate was approximately 60%, which was significantly lower than the initially assumed 70%. To ensure the statistical power and reliability of the study results, we recalculated the sample size based on the actual event rate. The new calculation indicated that, with an anticipated hazard ratio of 0.65, event rate of 60%, a significance level (α) of 0.5 and a power of 0.80, the minimum required sample size was calculated to be 80 patients. Considering a dropout rate of 20%, the final sample size was determined to be 96 patients. A total of 103 patients were included in this study, which met the sample size required for the study.

We have added relevant contents to the manuscript.

-Good to clarify the technique used for the assessment of microsatellite stability (IHC, PCR, both?)

Response: In our study protocol, immunohistochemistry was required to verify microsatellite stability.

-The title wording of section "2.3 Description of the results" should be amended as this section does not report results, but it describes the evaluation criteria for the study endpoints.

Response: We've amended it as "2.3 Outcomes" in the manuscript.

-History of previous surgery was an independent prognostic factor. Does this refer to the surgery of the primary tumour or metastases?

Response: We meant the surgery of the primary tumor, and we amended the wording in the manuscript. However, the content of multivariate analysis was deleted according to the opinion of Reviewer 3 (8. The results from the multivariable analysis show wide 95% CI, which is expected as the number of deaths is small. The estimated hazard ratios are unstable and I would refrain from reporting these results).

-There seems to be a substantial difference between mPFS (4.1 months) and mOS (14.1 months). This may suggest an impact of post-study treatments or a patient selection bias. The inclusion of a prognostically favourable population is suggested by the fact that none of the independent prognostic factors for PFS is independently prognostic for OS.

Response: As you said, we have observed that patients have a longer survival after disease progression than patients with conventional treatment. Since the majority (60.2%) of patients did not receive follow-up therapy, we believe that

the survival benefit was due to the combination of regorafenib and sintilimab, and we suspect that this may be related to the long-term effects of targeted therapy combined with immunotherapy, but the reasons for this need to be further explored.

-To clarify if the 20 tumour samples from the primary tumours were taken before study entry or at the time of diagnostic biopsy or initial resection.
Response: The tumor samples were taken at the time of diagnostic biopsy or initial resection. This has been revised in the manuscript.

-Not sure I understand the following sentence: “A comparison of tumorous paracancerous tissues revealed that the expression patterns of the upregulated genes in the treatment-sensitive group (DYRK2, ID4) were under expressed in the tumor tissues, and the genes that were downregulated in the treatment-sensitive group (TRMT112, ULBP1) were highly expressed in the tumor tissues (Figure 7-C)”.
Response: We changed the wording to make the meaning of the sentence easier to understand.

-Reference 8 appears inappropriate as it includes a study in hepatocellular carcinoma rather than colorectal cancer.
Response: Yes, the reference should be “FDA Approval Summary: Pembrolizumab for the First-line Treatment of Patients with MSI-H/dMMR Advanced Unresectable or Metastatic Colorectal Carcinoma”, and this has been revised in the manuscript.

-Recent and relevant studies should be discussed in the introduction and/or discussion section, such as the ATEZOTribe trial. Similarly to what the authors did, this study also evaluated the Immunoscore IC, taking into account not only the density of CD3/CD8 cells but also their spatial distribution and proximity to the tumour cells.
Response: Yes, our study explored the relationship between the distance between tumor cells and immune cells and the efficacy, and the relationship between IC score and the efficacy was explored in the ATEZOtribe study, which is worth our reference. We have already added relevant content to the Discussion section of the manuscript.

-Table 1 is reported twice with slighlty different headings. Please clarify.
Response: This has been revised in the manuscript.

-In Figure 1, 8 patients are indicated as being excluded but reasons for exclusions are reported only for 5 of them.
Response: "3 others" was missing due to incomplete word file display. This has been revised in the manuscript.

-Numbers at risk should be included in all figures

Response: This has been revised in the figures.

-Forest plots should report the 95%CI bars for all variables (including those which are used as reference).
Response: This has been revised in the figures.

Reviewer #2 (cancer therapy, clinical trial):

Liu R et al. presents a phase II study evaluating the efficacy and safety of the combination of regorafenib and sintilimab in patients with MSS mCRC. This study suggests promising efficacy with median OS of 14.1 months and 21.4% of ORR in patients who show typically poor response to immunotherapy.

1. The study reported a sample size of 82 patients but enrolled 103. Please clarify the reason for this discrepancy and assess the potential impact on the study's power.
Response: This study was an exploratory single-arm study, which was calculated using the one-sample log rank test. Based on previous literature and related research, it was assumed that the addition of sintilimab to regorafenib could elevate the median OS from 8.8 months to 13.5 months, with an anticipated hazard ratio of 0.65 and event rate of 70%. This assumption was used to calculate the sample size to ensure that, with a significance level (α) of 0.5 and a power of 0.80, the number of events was 48, the initially calculated sample size was 68 participants. Considering a dropout rate of 20%, the initially sample size was determined to be 82 patients.
However, during the actual study, we observed that the event rate was approximately 60%, which was significantly lower than the initially assumed 70%. To ensure the statistical power and reliability of the study results, we recalculated the sample size based on the actual event rate. The new calculation indicated that, with an anticipated hazard ratio of 0.65, event rate of 60%, a significance level (α) of 0.5 and a power of 0.80, the minimum required sample size was calculated to be 80 patients. Considering a dropout rate of 20%, the final sample size was determined to be 96 patients. A total of 103 patients were included in this study, which met the sample size required for the study.
We have added relevant contents to the manuscript.

2. The data seems to be that RAS/RAF status might be a prognostic rather than predictive factor. Moreover, the findings contradict those of the LEAP-017 trial. Please elaborate on this discrepancy and discuss its implications.
Response: The RAS/RAF grouping in the baseline analysis of the LEAP-017 trial was for all patients (including pembrolizumab+lenvatinib and SOC -Rego/TAS102), and the association between RAS/RAF stratification and efficacy was not analyzed in different cohorts of medication selection. Importantly, in the LEAP-017 trial

group, the survival benefit of RAS/RAF wild-type was lower than that of RAS/RAF mutant-type, and lower than that of SOC- regorafenib (Figure 1 below). It is reasonable to assume that the combined effect of lenvatinib with immunotherapy was limited due to its target.

Regorafenib combined with sintilimab was used in our study. Regorafenib had relatively clear combined immune targets, which improved the RAS/RAF wild-type survival benefits of patients enrolled in this study. In addition, although the OS of the RAS/RAF mutant-type population in this study was lower than that of the RAS/RAF wild-type population, the survival benefit of RAS/RAF mutant-type was very similar to the efficacy of regorafenib with immunotherapy that has been reported (Figure 2 below).

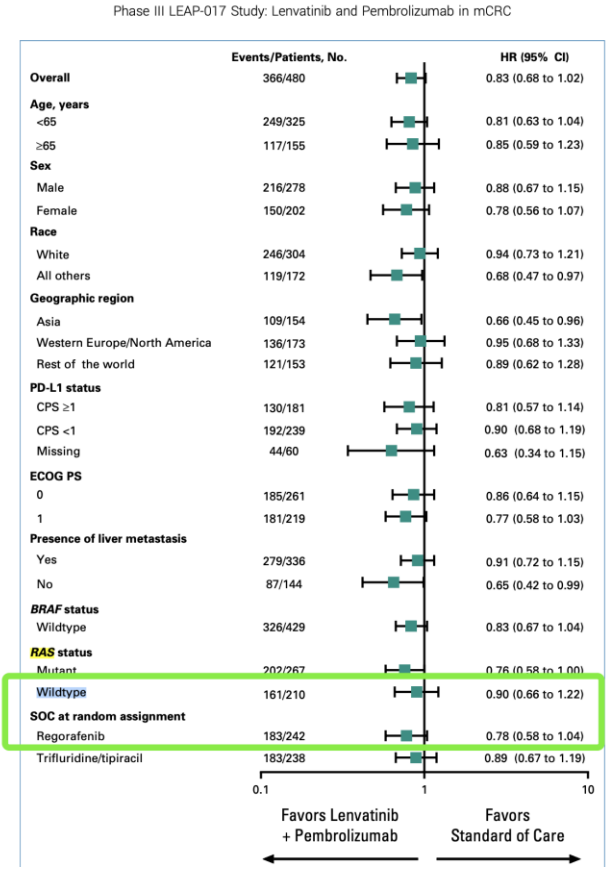

Figure 1. Subgroup analysis in LEAP-017 study.

	Fukuoka REGONIVO (N=25)[1†]	Kim REGONIVO (N=52)[2]	Fakih REGONIVO (N=70)[3]	Cousin REGOMUNE (N=47)[4]	Wang REGOTORI (N=39)[5‡]	Grothey CORRECT (N=760)[6]	Eng IMBlaze 370 (N=90)[7]	Gomez-Roca LEAP-005 pembro+lenva (N= 32)[8§]
中位年龄, 岁(范围)	55 (31, 77)	56 (31, 79)	57 (34, 85)	62 (26, 83)	53 (37, 69)	61 (54, 67)	59 (52, 66)	56 (36, 77)
男性, %	72	54	59	74	51	62	57	81
ECOG PS 0/1, %	100/0	38/62	51/49	60/40	8/92	52/48	50/50	-/56
右半/左半肿瘤, %	20/80	58/42	36 (右半结肠) /64 (左半结肠/直肠)	-	-	-	30/49	-
*(K)RAS*突变, %	24	71	61	64	51	54	54/46	-
转移部位, %								
淋巴结	60	–	-	30	46	-	-	-
肝	52	73	67	74	69	-	66	-
肺	64	63	73	79	64	-	-	-
腹膜	16	39	-	32	26	-	-	-
既往治疗, %								
2线	20	58	43	30	67	27	77	91
≥3线	80	42	53	57	33	73	23	3
疗效								
中位OS, 月(95% CI)	NR (9.8, NR)	11.1 (7.7, 14.5)	11.9 (7.0, NE)	10.8 (5.9, NR)	15.5 (10.3, NR)	6.4 (IQR 3.6–11.8)	8.51 (6.41, 10.71)	7.5 (3.9, NR)
中位PFS, 月(95% CI)	7.9 (2.9, NR)	4.3 (1.6, 7.0)	1.8 (1.8, 2.4)	3.6 (1.8, 5.4)	2.6 (2.0, 4.3)	1.9 (IQR 1.6–3.9)	2.00 (1.87, 3.61)	2.3 (2.0, 5.2)
ORR, %	36	n=40 13**	7	n=43 0	n=33 15	1	2 (95% CI 0.3, 7.8)	22
DCR, %	84	n=40 63	39	n=43 53	n=33 36	41	34	47

*各研究数据比较仅用于说明目的。研究可能在人群、设计和终点方面有所不同，因此不应进行直接比较；†数据来自mCRC患者队列; ‡所提供的数据来自接受瑞戈非尼80mg患者队列; §数据来自mCRC患者队列; ¶除非另有说明，数据包括20例mCRC患者和15例胃癌患者; **1例患者未证实部分缓解。
apat, 阿帕替尼; camre,卡瑞利珠单抗; CI, 置信区间; DCR,疾病控制率; ECOG PS,美国东部肿瘤合作组体能状态; IQR,四分位差; lenva, 仑伐替尼; mCRC, 转移性结直肠癌; NE,不可评估(由于数据删失); NR, 未报告; OS, 总生存期; pembro, 帕博利珠单抗; PFS, 无进展生存期。
1. Fukuoka S, et al. *J Clin Oncol* 2020;38:2053–2061; 2. Kim RD, et al. [Poster due to be presented at WCGI 2021]; 3. Fakih M, et al. *J Clin Oncol* 2021;39:3560; 4. Cousin S, et al. *Clin Cancer Res* 2021;27:2139–2147; 5. Wang F, et al. *Ann Oncol* 2020;31(Suppl 4):S425; 6. Grothey A, Van Cutsem E, et al. *Lancet* 2013;381:303–312; 7. Eng C, et al. *Lancet Oncol* 2019;20:849–861; 8. Gomez-Roca C, et al. *J Clin Oncol* 2021;3(Suppl):94; 9. Xiao L et al. ESMO 2020, 442P.

Figure 2. Research overview of regorafenib and other TKI combined immunotherapy for mCRC

3. Please specify whether the reported ORR is confirmed or unconfirmed.

Response: Yes, the ORR was based on confirmed response, and this has been revised in the Methods section of the manuscript.

4. Figure 5A shows a case classified as PR with less than 30% tumor shrinkage.

Response: Sorry for our oversight. Compared with baseline, the percentage change in the sum of target lesion diameter was -30.61%, and the -26.53% shown in our figure is not the best effect, but the first evaluation result. We modified the figure accordingly.

住院号	最佳评效	基线靶病灶直径	最佳评效直径	缓解程度	缓解程度1
538196	PR	98	68	-30.61%	-26.53%

5. Please analyze and report efficacy according to PD-L1 expression levels (e.g., CPS or TPS).

Response: We enrolled patients from November 2020 to February 2023. During this period, there was no unified standard for PD-L1 expression detection scheme in China, so the expression of PD-L1 was not required in our research scheme. On the other hand, the predictive effect of PD-L1 expression on immunotherapy efficacy in mCRC patients remains unclear, so PD-L1 expression testing is not routinely performed in mCRC patients. So we are sorry we can't get this part of the results.

6. Please provide detailed information on serious adverse events (SAEs) to comprehensively assess the safety profile of the combination therapy.

Response: Treatment was generally well tolerated and there was no SAE in our study.

7. The authors claim that Chinese patients may benefit more from the combination therapy. Please provide evidence to support this assertion and discuss potential underlying mechanisms.

Response: Compared with CORRECT study (mPFS=1.9m, mOS=6.4m), CONCUR study results in Asian population seem to be better (whole population: mPFS=3.2m, mOS=8.8m, Chinese population: mPFS=2.0m, mOS=8.4m) and the incidence of adverse events of 3 and above was lower. A recent "Multi-center Retrospective real-world study in China" from the National Cancer Center of China (Regorafenib monotherapy or combined with an immune-checkpoint inhibitor as later-line treatment for metastatic colorectal cancer: a multicenter, real-world retrospective study in China. BMC Cancer. 2024;24(1):22) analyzed the efficacy of regimen containing regorafenib as third-line or later-line treatment for mCRC in ten hospitals in China from August 2017 to June 2020. The results showed that combined immunotherapy with regorafenib significantly improved the clinical outcome of mCRC compared to regorafenib alone: OS (13.5 vs. 10.0 months, p = 0.001).

The potential underlying mechanisms are unclear up to now. Immune microenvironment analysis in the REGOMUNE study suggested there were associations between macrophage infiltration and a poor PFS/ OS, and CD8+ T-cell infiltration and improved outcomes. Regorafenib significantly increased CD8+T cell infiltration in 60% of patients, and PFS and OS in these patients were significantly better than those in patients with no change in CD8+T cell infiltration. Our study revealed that patients in the sensitive group not only exhibited specific gene expressions regulating tumor metabolic pathways but also have a significantly higher density of CD8+T cells compared to those in the treatment-resistant group, with the average distance between CD8+T cells and tumor cells also being shorter in the treatment-sensitive group. The exact mechanism of regorafenib combined with immunotherapy still needs to be further explored.

8. Given the relevance of the LEAP-017 trial, a detailed comparison of the results and implications is warranted.

Response: We compared the main end-points, liver metastasis status, RAS/RAF status of our study and LEAP-017 study, and discussed the similarities and differences in results and possible mechanisms. We have added relevant information in the Discussion section of the manuscript.

Minor

1. Cite the CONCUR trial as a reference for the assessment criteria threshold.

Response: We've cited the CONCUR trial as a reference in the manuscript.

2. Ensure consistent use of decimal points throughout the manuscript.

Response: We've kept one decimal place uniformly in the manuscript.

3. In conclusion, revise "particularly in patients with RAS wild-type gene" to "particularly in patients with RAS/RAF wild-type gene."
Response: This has been revised in the manuscript.

Reviewer #3 (Biostatistician, clinical trial):

In this phase II trial the authors report the efficacy and safety results from regorafenib plus sintilimab as salvage treatment for microsatellite stable (MSS) in mCRC patients. There are several a few flaws with this trial, mainly the choice of the primary endpoint and the incorrect use of the test statistic to determine the sample size. It would be helpful to see the protocol.

1) Primary Endpoint of OS in a Single-Arm Phase II Trial: Selecting overall survival (OS) as the primary endpoint in a single-arm Phase II trial is unconventional. Typically, OS is reserved for randomized Phase III trials, as comparing survival outcomes requires a control arm. The authors need to clearly justify the choice of OS and explain how they plan to interpret this endpoint in the absence of a comparator group.
Response: PFS is a common endpoint in Phase II oncology clinical trials because it quickly reflects the effect of treatment on disease control and is not affected by subsequent treatment. But over the course of our study, there was an outbreak of covid-19, and PFS as the primary endpoint could have been affected by evaluation bias, especially when imaging assessments were not rigorously blind or independently evaluated. OS is a more objective and direct endpoint that is not subject to the subjectivity of imaging assessment.
We selected OS as the primary endpoint primarily because of its clear clinical relevance. Overall survival is the most intuitive end point and can directly reflect the survival benefit of patients.
In the absence of a control group, we plan to interpret the OS by comparing it with historical control data. We will use the established median OS from previous studies as reference criteria and compare the OS observed in our study with these criteria to evaluate the potential effect of additional treatments.
In addition to OS, we selected other secondary endpoints (such as PFS, ORR, etc.) that could complement the OS results and help explain the overall effect of treatment. By evaluating efficacy from multiple angles, we hope to provide a more comprehensive picture of treatment outcomes.

2) It is not clear how the sample was computed in this single arm phase II trial. The authors refer to the single log rank test. This reviewer is not familiar with this test as the log-rank test Can they elaborate? What was the alternative hypothesis?
Response: This study was an exploratory single-arm study, which was calculated using the one-sample log rank test. Based on previous literature and related

research, it was assumed that the addition of sintilimab to regorafenib could elevate the median OS from 8.8 months to 13.5 months, with an anticipated hazard ratio of 0.65 and event rate of 70%. This assumption was used to calculate the sample size to ensure that, with a significance level (α) of 0.5 and a power of 0.80, the number of events was 48, the initially calculated sample size was 68 participants. Considering a dropout rate of 20%, the initially sample size was determined to be 82 patients.

However, during the actual study, we observed that the event rate was approximately 60%, which was significantly lower than the initially assumed 70%. To ensure the statistical power and reliability of the study results, we recalculated the sample size based on the actual event rate. The new calculation indicated that, with an anticipated hazard ratio of 0.65, event rate of 60%, a significance level (α) of 0.5 and a power of 0.80, the minimum required sample size was calculated to be 80 patients. Considering a dropout rate of 20%, the final sample size was determined to be 96 patients. A total of 103 patients were included in this study, which met the sample size required for the study.

We have added relevant contents to the manuscript.

3) In Table 1, can you replace ECOG Score with "ECOG PS" and also uncouple ECOG PS as 0, 1, to present these two groups separately? Also, where there are several possible answer, indicate that the categories are non-mutually exclusive (such as prior previous treatment)

Response: This has been revised in Table 1 as you advised.

4) For the ORR, were the PR and CR based on confirmed response?

Response: Yes, the PR and CR were based on confirmed response, and this has been revised in the Methods section of the manuscript.

5) How many deaths were observed? Was the 55 deaths due to treatment or disease progression?

Response: Among the 103 patients, 55 patients had died until June 7, 2023. Fifty-three deaths were due to disease progression, one due to heart disease, and one due to thrombotic events. None was due to the treatment.

6) For all the 95% CI, please report 1 digit after the decimal point.

Response: This has been revised in the manuscript.

7) In Table 2, can you present grade 3, 4, and 5 as individual categories?

Response: As we described in the results section "There were no grade 4-5 TRAEs", so we changed "≥Grade 3 " to "Grade 3" in Table 2.

8) The results from the multivariable analysis show wide 95% CI, which is expected as the number of deaths is small. The estimated hazard ratios are unstable and I would refrain from reporting these results.

Response: We have deleted the multivariable analysis in the manuscript.

9) For the Kaplan-Meier curves (2A, 3A-3B, 4), label the x axis properly. Is it time since enrollment (months?)
Response: Yes, the x axis is the time since enrollment. This has been revised in the figures.

10) Remove the p-values from Figures 2B & 4B; as no prior hypotheses were specified and it would be hard to interpret the results.
Response: We've removed the p-values from Figures 2B & 4B as you advised.

11) What is the purpose of Figure 5A?
Response: Figure 5A showed the best percentage change in target lesion size from baseline and was a visual demonstration of ORR.

12) The resolutions of some of the figures are weak (such as Figure 7).
Response: This has been revised in the figures.

13) Please check for typos: Trail should be trial, etc.
Response: This has been revised in the manuscript.

REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

The manuscript has been revised approximately. Please add the manuscript the following contexts.

1. There was no SAE in the safety section.

2. There was no data regarding PDL1 expression as a limitation in discussion section.

Reviewer #3 (Remarks to the Author):

The authors have addressed my comments, except that they need to specify if the type I error rate was one-sided or two sided.

REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

The manuscript has been revised approximately. Please add the manuscript the following contexts.

1. There was no SAE in the safety section.

Response: Thank you for your suggestion. Your comments have been extremely helpful, and this has been added in the manuscript.

2. There was no data regarding PDL1 expression as a limitation in discussion section.

Response: Thank you very much for your valuable suggestions. This has been added as a limitation in discussion section in the manuscript

Reviewer #3 (Remarks to the Author):

The authors have addressed my comments, except that they need to specify if the type I error rate was one-sided or two sided.

Response: Thank you for your valuable suggestion. Type I error rate was two-sided, and this has been added in the manuscript.