Statistical review of *Favipiravir versus Arbidol for COVID-19: A Randomized Clinical Trial*

March 31, 2020

Dr Jack Wilkinson, PhD - Research Fellow, Centre for Biostatistics, Manchester Academic Health Science Centre, University of Manchester.

Dr Darren Dahly, PhD - Principal Statistician, HRB Clinical Research Facility - Cork. HRB Trials Methodology Research Network. Senior Lecturer in Patient Focused Research Methods, University College Cork School of Public Health

The following review has been prepared in collaboration with members of the MRC-NIHR Trials Methodology Research Partnership [1]. The reviewers named above, and other, unnamed discussants of the paper, are all qualified statisticians with experience in clinical trials. Our objective is to provide a rapid review of publications, preprints and protocols from clinical trials of COVID-19 treatments, independent of journal specific review processes. We aim to provide timely, constructive, focused, clear advice aimed at improving both the research outputs under review, as well as future studies. Given our collective expertise (clinical trial statistics) our reviews focus on the designs of the trials and other statistical content (methods, presentation and accuracy of results, inferences).

DOI: 10.5281/zenodo.3734198

Study Summary

Here we review *Favipiravir versus Arbidol for COVID-19: A Randomized Clinical Trial,* by Chen *et al* [3]. The paper reports a multicenter, open-label, randomized superiority trial that took place in three hospitals in China. The study participants were 240 patients with COVID-19 pneumonia. These patients were randomized in equal numbers to two groups: the first received the antiviral drug favipiravir as an experimental treatment (n = 120; 1600 mg/time on the first day, twice a day; 600 mg/time from the second day to the end of the experiment, twice a day), while the second received arbidol (another antiviral drug that is already a part of standard care in China; n = 120; 200 mg each time, 3 times a day, from the first day to the end of the trial). The inclusion criteria were patients diagnosed with COVID-19 within 12 days of onset of symptoms, aged 18 years or older, giving informed consent. The authors excluded patients with known allergies to either treatment, highly elevated ALT/AST ratio, Child-Pugh Score of C, critical patients with expected survival < 48 hours, women with a positive pregnancy test, people with HIV infection,

and those who were otherwise “considered unsuitable”. The primary outcome was clinical recovery at +7 days post-randomization, where clinical recovery was defined as the return of fever, respiratory rate, oxygen saturation and coughing to normal levels for at least 72 hours. Four patients in the active arm were lost to follow-up.

The authors reported that among all patients, 61% (71/116) of those in the favipiravir arm clinically recovered by day +7, compared to 52% (62/120) of those in the arbidol arm (RD = 9.5%; p = 0.14). However, when only considering the patients with less severe illness (“ordinary patients”), 71% (70/98) of those in the favipiravir arm clinically recovered by day +7, compared to 56% (62/111) of those in the arbidol arm (RD = 15.5%; p = 0.02). Based on these findings, they concluded that favipiravir led to a better chance of clinical recovery at +7 days than arbidol in ordinary COVID-19 patients previously untreated with antivirals.

We sincerely thank the authors for their contribution to our collective understanding of COVID-19, for their commitment to the timely dissemination of research results, and for their transparency in sharing their full protocol and registering their trial.

Major comments

The main finding is in a subgroup of patients that wasn’t prespecified in the protocol.

While the patients in the favipiravir arm were more likely to have clinically recovered by day +7 than patients in the arbidol arm (a risk difference of 9.5%), this observed difference could have been plausibly explained by sampling error (i.e. this observation wouldn’t have been very unusual even if there was no difference between the two arms; p = 0.14). In the discussion, the authors attribute this “non-significant” finding to the fact that the proportion of “critical” patients, with poorer outcomes overall, was twice as high in the favipiravir arm compared to the arbidol group, and thus helped obscure any benefit of favipiravir over arbidol. Based on this, the authors decided to exclude the critical patients (18 from the favipiravir and 9 from the arbidol arm), and estimate the treatment effect in the remaining sample of “ordinary” patients. In this subgroup, they found a risk difference of RD = 15.5% (p = 0.02) . The authors’ overall conclusions are largely driven by this later finding.

Unfortunately, this subgroup analysis wasn’t specified in the study protocol or trial registry. Further, while the protocol does describe four clinical types (mild, moderate, severe and critical), these don’t exactly match up to the “ordinary” and “critical” patient types reported in the paper (mild and moderate were seemingly combined to define ordinary; the definition of severe in the protocol matches the definition of critical in the paper). Finally, because the statistical analysis plan in the protocol doesn’t specify how these subgroups would be used in the analysis, it is

unclear whether other subgroup analyses were performed but not reported. This overall flexibility in the analysis, as well as the lack of transparency, compel us to downgrade the strength of this finding [4].

Recommendations:
For future studies

- Include a detailed statistical analysis plan in your protocol; ensure your protocol and statistical analysis plan are registered and available prior to starting patient recruitment; ensure that the analysis reported in the trial's paper matches the pre-specified analysis plan.

For this study

- The conclusions should be largely based on the pre-specified primary analysis in the entire sample.
- The authors should more clearly explain how these decisions regarding subgroups were made, particular if there were any unreported analyses.

For the reader

- While the above concerns don't fully invalidate the finding for a beneficial effect of favipiravir in the subgroup of "ordinary" patients, they should be interpreted with caution.

Details of the randomization procedure were lacking, and there was no allocation concealment.

The implementation of the randomisation is not clearly described, either in the preprint or protocol. Further, it is unclear whether or not the allocation was prospectively concealed. While we randomize to prevent biases in how patients are allocated to arms, this can be undermined where there is no allocation concealment (meaning that allocations aren't made until patients are unambiguously consented and enrolled onto the trial; and that allocations can't be altered once made [5]). For example, a clinician who knows (or can make an informed guess) about the next allocation in the sequence might be more likely to enrol a patient with more severe symptoms if they know they will go into the arm receiving the experimental treatment.

Recommendations:
For future studies

- Please employ a rigorous allocation concealment process, preferably one that is electronic and external to the study team. Whatever process is used, ensure that its fidelity can be evaluated by study monitors.

For this study

- Please clarify any additional details about how the randomization was conducted, and how allocation concealment was achieved (if at all), per CONSORT guidelines (see below).

There was no blinding.

To prevent bias, it is important that patients and investigators are unaware of what study arm a given patient is in. In this case, blinding might have been complicated by the fact that favipiravir and arbidol have different treatment regimens, although the *double dummy* method might have been possible. We are of course aware that the challenging circumstances of the trial might have also prevented proper blinding, but it is still important that any results are interpreted in light of this limitation.

Recommendations:
For future studies

- Whenever possible, use blinding procedures to prevent post-randomization biases.

For this study

- Please interpret the results in light of fact that there was no blinding.

The data were analysed with suboptimal statistical tests/models.

There were a number of issues regarding the analyses that could be improved upon.

- For the primary outcome, it was not clear what method was used to arrive at the reported 95% confidence intervals or p-values.

- The authors report a risk difference for the primary outcome. However, the risk difference as a measure of treatment effect can be problematic because we know it can't be constant across levels of baseline risk, i.e. it can't be additive. It would be better, and consistent with CONSORT guidelines, to also report a relative measure of the effect size; preferably an odds ratio, since it can be constant across levels of baseline risk.

- For their primary outcome, the authors essentially dichotomized a time-to-event measure (days until clinical recovery) but they did not do this for individual components of the primary outcome, such as the number of days until fever returned to normal levels. Using a time-to-event version for the primary outcome (i.e. days until clinical recovery) would have likely resulted in a more efficient use of the data (e.g. narrower confidence intervals; more power to detect a given effect).

- The analyses for all reported end points used crude tests/models, such as log-rank tests for time to event outcomes. However, there were a number of factors measured at baseline (prior to randomization) that were likely prognostic for outcomes (e.g. age, sex, smoking hisotry, diease severity, viral load). The authors thus should consider covariate-adjusted estimates of treatment effects [6] using the appropriate multivariable

statistical model (logistic regression for binary outcomes; Cox proportional hazards models for time-to-event outcomes). This would result in more appropriate conditional measures of treatment effects, as well as more efficient estimates, with narrower confidence intervals, given the fixed sample size. For the primary analysis, it is crucial that these covariates are prespecified, and not be selected on the basis of observed "imbalances" in the data.

- Use of multivariable models would also allow for the proper consideration of heterogeneity of treatment effects across subgroups. For example, the authors infer that there is a beneficial effect of favipiravir in the group of "ordinary" patients, but not in the "critical" patients. This inference is based on the finding of a "significant" result in the first group, and a non-significant result in the latter group. However, differences in significance should not be used to imply significant differences [7]. It would be more appropriate to test or model potential subgroup-specific effects using an interaction term in the multivariable model [8].

Recommendations:

For future studies

- Whenever possible, please consult or collaborate with an experienced trial statistician, especially at the design phase.

For this study

- Report additional post-hoc analyses following the advice above.

Minor points

- How subjective is this outcome assessment? It seems well defined, but perhaps there is some subjectively there. This is relevant given that the study wasn't blinded.

- There were two secondary outcomes listed in the protocol but not reported in the trial paper (PCR, and ICU admission).

- It would assist efforts to synthesise data across trials if authors reported standardised outcomes as detailed in the appropriate core outcome set. There are several complementary core outcome sets for COVID-19, and authors might consider which set is most appropriate: http://www.comet-initiative.org/Studies/Details/1538 (note: these outcome sets are still under development, and require thoughtful consideration before adopting).

- This passage, *"The cases of respiratory failure in the two group were both 4",* didn't match the figures in the table, which indicated one in the favipiravir group and four in the arbidol group.

- The approach to missing data as described in the protocol is unclear. The authors note that “The lack of primary efficacy indicators was filled using principles that were not conducive to the experimental group. The absence of secondary efficacy indicators and safety indicators were not filled.” Thus it does not appear that any missing data strategy has been employed, though four participants “withdrew consent” in the favipiravir group post-randomisation and were excluded from the analysis. Other analysis sets were described in the protocol, but not presented in the manuscript.

- Authors suggest that the choice of comparator was pragmatic, and that the efficacy of the comparator in this context is not known, i.e. we can't be certain the trial has assay sensitivity [9]. Thus, while favipiravir might have outperformed arbidol, it is possible that neither is beneficial.

- Analyses have also been performed in the subgroup of patients with diabetes or hypertension. These also appear to be post-hoc, as they are not described in the protocol or trial registry.

- It was also unclear whether the patient types were a) determined pre-randomisation and b) fixed, such that a participant could not transition from “ordinary” to “critical” post-randomisation.

- Fever and cough duration were analysed in subgroups determined by “ordinary” status and presence of those symptoms are a part of that definition.

- The study takes a frequentist approach to making statistical inferences, but without any consideration of error controls (e.g. consideration of multiplicity) in their conclusions.
`

Open Data

None provided – authors state it is available on request

Open Analysis Code

No.

Pre-registered study design

A registration was made several days after the start of the study (ChiCTR2000030254).

PubPeer

There may be comments on the PubPeer page for the published version of this paper.
https://pubpeer.com/publications/9C9A1AA1343F05CB5458B573C3694B

References

1. MRC-NIHR Trials Methodology Research Partnership. https://www.methodologyhubs.mrc.ac.uk/about/tmrp/
2. Creative Commons Attribution 4.0 International License. https://creativecommons.org/licenses/by-sa/4.0/
3. Chen, C. *et al. Favipiravir versus Arbidol for COVID-19: A Randomized Clinical Trial*. http://medrxiv.org/lookup/doi/10.1101/2020.03.17.20037432 (2020) doi:10.1101/2020.03.17.20037432.
4. Munafò, M. R. *et al.* A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 0021 (2017).
5. Vickers, A. J. How to randomize. *J. Soc. Integr. Oncol.* **4**, 194–198 (2006).
6. Kahan, B. C., Jairath, V., Doré, C. J. & Morris, T. P. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials* **15**, 139 (2014).
7. Gelman, A. & Stern, H. The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. *Am. Stat.* **60**, 328–331 (2006).
8. Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J. & Drazen, J. M. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *N. Engl. J. Med.* **357**, 2189–2194 (2007).
9. *ICH HARMONISED TRIPARTITE GUIDELINE Choice of control group and related issues in clinical trials (E10)*. (2000).
10. Schulz, K. F., Altman, D. G., Moher, D. & for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c332–c332 (2010).

CONSORT CHECKLIST

To support the review, we completed the CONSORT checklist [10] below. Material taken directly from the paper (or trial registry) is in *italics.* Our additional comments are in **bold**.

Title and abstract

1a Identification as a randomised trial in the title

Favipiravir versus Arbidol for COVID-19: A Randomized Clinical Trial

1b Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)

Title: Identification of the study as randomised	Yes
Authors: Contact details for the corresponding author	Yes
Trial design: Description of the trial design (eg, parallel, cluster, non-inferiority)	Yes
Methods	
Participants: Eligibility criteria for participants and the settings where the data were collected	YES
Interventions: Interventions intended for each group	YES
Objective: Specific objective or hypothesis	YES
Outcome: Clearly defined primary outcome for this report	YES
Randomisation: How participants were allocated to interventions	NA
Blinding (masking): Whether or not participants, care givers, and those assessing the outcomes were blinded to group assignment	NA
Results	
Numbers randomised: Number of participants randomised to each group	YES
Recruitment: Trial status	NO
Numbers analysed: Number of participants analysed in each group	YES
Outcome: For the primary outcome, a result for each group and the estimated effect size and its precision	YES
Harms: Important adverse events or side-effects	YES
Conclusions: General interpretation of the results	YES

Trial registration: Registration number and name of trial register	YES
Funding: Source of funding	NO

Introduction

Background and objectives

2a Scientific background and explanation of rationale

Yes.

2b Specific objectives or hypotheses

Objective: To compare the efficacy and safety of favipiravir and arbidol to treat COVID-19 patients on 7 day's clinical recovery rate.

In this study, we hypothesized that favipiravir would be non-inferior to arbidol in terms of efficacy for moderate symptoms, and improves outcomes clinical recovery of fever, cough, and breathing difficulties compared with antiviral efficacy of arbidol. We therefore assessed the clinical efficacy and safety of favipiravir versus arbidol as treatment for SARS-CoV-2.

Methods

Trial design

3a Description of trial design (such as parallel, factorial) including allocation ratio

We conducted a prospective, multicenter, open-labelled, randomized superiority trial in 240 patients with COVID-19 pneumonia at three hospitals (120 patients from Zhongnan Hospital of Wuhan University, 88 patients from Leishenshan Hospital, 32 patients from The Third People's Hospital of Hubei Province).

In this study, according to the proportion of 1:1 between the experimental group (favipiravir) and the control group (arbidol)

3b Important changes to methods after trial commencement (such as eligibility criteria), with reasons

Not reported.

Participants

4a Eligibility criteria for participants

Patients were eligible if they met all the following criteria: (1) aged 18 years or older; (2) voluntarily signed informed consent; (3) the initial symptoms were within 12 days; (4) diagnosed as COVID-19 pneumonia.

Patients meeting any of the following criteria were excluded: (1) were allergic to fabiravir or arbidol; (2) ALT/AST increased 5 times higher than the upper limit of normal, or with child Pugh C; (3) critical patients whose expected survival time < 48 hours; (4) childbearing age women with positive pregnancy test; (5) with HIV infection; (6) were considered unsuitable by researchers.

From the registry (ChiCTR2000030254):
Inclusion criteria
(1) Aged 18 years or older; (2) Voluntarily signed informed consent; (3) Hospitalized patients diagnosed as COVID-19.
Exclusion criteria
(1) Allergic to fabiravir or abidol; (2) ALT/AST increased 5 times higher than the upper limit of normal, or with child Pugh C; (3) Severe patients with expected survival time < 48 hours; (4) Pregnancy; (5) HIV positive; (6) Considered unsuitable by researchers.

4b Settings and locations where the data were collected

...(120 patients from Zhongnan Hospital of Wuhan University, 88 patients from Leishenshan Hospital, 32 patients from The Third People's Hospital of Hubei Province). Patients were prospectively enrolled and followed-up from Feb 20, 2020 to Mar 12, 2020.

Interventions

5 The interventions for each group with sufficient details to allow replication, including how and when they were actually administered

The experimental group (famiravir) was treated with routine treatment + famiravir tablets (1600 mg/time on the first day, twice a day; 600 mg/time from the second day to the end of the experiment, twice a day).

The control group (arbidol) was treated with routine therapy + arbidol (200 mg each time, 3 times a day, from the first day to the end of the trial).

The course of treatment in both groups was 7-10 days. If necessary, the treatment time could be extended to 10 days according to the judgment of researchers.

Outcomes

6a Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed

The primary outcome was the clinical recovery rate at 7 days or the end of treatment, which was stratified as ordinary patients with COVID-19, critical patients with COVID-19, COVID-19 patients with hypertension and/or diabetes **(see next paragraph for definitions of these strata)**. *The recovery of fever, respiratory rate, oxygen saturation and cough relief after treatment were defined as clinical recovery, and the recovery state lasted no less than 72 hours. It needs to meet several conditions: axillary temperature ≤ 36.6 °C; respiratory frequency ≤ 24 times/min; Oxygen saturation ≥ 98% without oxygen inhalation; mild or no cough. The armpit temperature, respiratory rate, oxygen saturation without oxygen, oxygen therapy and noninvasive positive pressure ventilation (NPPV) were recorded in daily follow-up. Repeated measurements were made at least twice in each follow-up. The measurements were taken after 15 minutes rest at room temperature (23±2 °C).*

Classification criteria of ordinary COVID-19 patients and critical COVID-19 patients: (1) Ordinary COVID-19 patients: has a fever, respiratory symptom, can be observed by imageology methods. (2) Critical COVID-19 patients: meeting any of the following case: a. dyspnea, RR > 30 times/min; b. the SpO2 < 93% in the resting state; c. PaO2/FiO2 < 300mmHg (1 mmHg = 0.133 kPa). PaO2/FiO2 should be corrected according to the formula: PaO2/FiO2 × [atmospheric pressure (mmHg)/760]. The pulmonary imaging showed that the lesions progressed more than 50% within 24-48 hours, and the patients were classified as critical patients.

Secondary outcomes included the time from randomization to fever reduction (patients with fever at the time of enrollment), the time from randomization to cough relief (patients with moderate or severe cough at the time of enrollment), the rate of auxiliary oxygen therapy or noninvasive mechanical ventilation during the trial, the all-cause mortality during the trial, the rate of respiratory failure during the trial (defined as SPO2 ≤ 90% or PaO2/FiO2 < 300 mmHg without oxygen inhalation, and requires oxygen therapy or higher respiratory support). Blood biochemistry, urine routine, coagulation function, C-reactive protein, nucleic acid and CT were examined on the third day (D3±1 day) and the seventh day (D7±1 day) after taking the drug, and the adverse events and concomitant medication were observed.

From the registry:
Primary:

- *Clinical recovery rate of day 7*

Secondary:

- *All-cause mortality during the trial*
- *The rate of respiratory failure during the trial*
- *The time from randomization to fever reduction*
- *The time from randomization to cough relief*
- *The time from randomization to dyspnea relief* **(Not time to event in paper)**
- *The rate of auxiliary oxygen therapy or noninvasive mechanical ventilation during the trial*
- *After one week of treatment, the negative rate of 2019-nCOV RT PCR test for upper respiratory tract specimens* **(Dropped from paper)**
- *The rate of ICU admission during the trial* **(Dropped from paper)**
- *Incidence of serious adverse events (SAE) during the trial*

6b Any changes to trial outcomes after the trial commenced, with reasons

None reported.

Sample size

7a How sample size was determined

Sample size estimation: the expected clinical recovery rate of the experimental group is 70%, the clinical recovery of the control group is 50%, α = 0.025 (single side), β = 0.20, power = 0.80. According to the distribution ratio of 1:1 between the experimental group and the control group, the statistical sample size is 92 participants in each group. The sample size increased about 20% considering factors such as shedding/elimination. The trial was designed to include 240 participants in the group, including 120 in the experimental group and 120 in the control group.

Confirmed.

```
>   library(pwr) # http://www.statmethods.net/stats/power.html
>   h1 <- (2 * asin(sqrt(.70))) - (2 * asin(sqrt(.50)))
>   pwr.2p.test(h = h1, sig.level = .05, power = 0.8)

     Difference of proportion power calculation for binomial distr
ibution (arcsine transformation)

              h = 0.4115168
              n = 92.69608
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: same sample sizes

> 92/0.8
[1] 115
>
```

7b When applicable, explanation of any interim analyses and stopping guidelines

Not applicable.

Randomisation

Sequence generation

8a Method used to generate the random allocation sequence

...randomized open label was produced by professional statistical software SAS9.4

8b Type of randomisation; details of any restriction (such as blocking and block size)

No additional information provided.

Allocation concealment mechanism

9 Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned

No information provided.

Implementation

10 Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions

From the protocol:
The specific random process was as follows: after confirming that the subjects provide informed consent, and meet all the inclusion criteria and does not meet any of the exclusion criteria, the pre-generated randomization table was used to obtain the subject's random number and drug allocation information.

Once the inclusion and exclusion criteria are confirmed, randomization should be performed as soon as possible (12 hours). According to the recorded inclusion criteria, the maximum acceptable time is 24 hours.

Blinding

There was no blinding.

Statistical methods

12a Statistical methods used to compare groups for primary and secondary outcomes

SAS9.4 software was used for statistical analysis.

For the main efficacy indicator/primary outcome (clinical recovery rate after 7 days or the end of treatment), the comparison between the experimental group and the control group adopts the optimal test. We calculated the bilateral 95% CI of the difference between the clinical recovery rate of the experimental group and the control group. If the lower limit was > 0, it was considered the experimental group (favipiravir) is superior to the control group (arbidol).

Log rank test was used to compare the “time” between the two groups.

For the secondary efficacy indicators/secondary outcomes, t test or Wilcoxon rank sum test (if t-test was not applicable) was performed for safety indicators and continuous variables, Wilcoxon rank sum test was used for grade variables.

Frequency or composition (%) were used for statistical description of classification indexes, and Chi-square test test or Fisher’s exact test was used for comparison between groups.

For all statistical tests, P value < 0.05 (bilateral) were considered as statistically significant.

12b Methods for additional analyses, such as subgroup analyses and adjusted analyses

Not applicable.

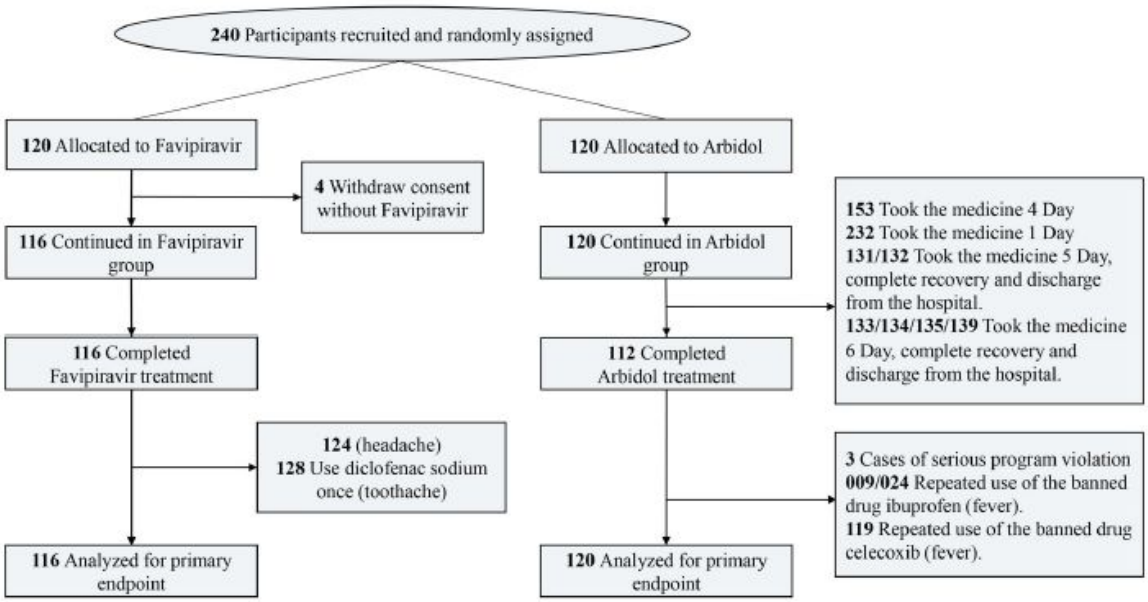
Results

Participant flow (a diagram is strongly recommended)

13a For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome

Total 236 patients with COVID-19 were enrolled in the full analysis set (FAS), 116 in the experimental group (favipiravir) and 120 in the control group (arbidol).

13b For each group, losses and exclusions after randomisation, together with reasons



Recruitment

14a Dates defining the periods of recruitment and follow-up

Patients were prospectively enrolled and followed-up from Feb 20, 2020 to Mar 12, 2020.

14b Why the trial ended or was stopped

It hit its recruitment target.

Baseline data

15 A table showing baseline demographic and clinical characteristics for each group

The characteristics of patients in the 2 groups were shown in table 1. In the experimental group, 59 were males and 57 were females, 87 (75.00%) were < 65 years and 29 (25.00%) were ≥ 65 years, 36 (31.03%) were with hypertension and 14 (12.07%) with diabetes. In the control group, 51 were males and 69 were females, 79 (65.83%) were < 65 years and 41 (34.17%) were ≥ 65 years, 30 (25.00%) were with hypertension and 13 (10.83%) with diabetes.

Table 1. Basic characteristics of the participants.

Variables	Favipiravir group (N = 116)	Arbidol group (N = 120)	P value
Gender			
Female, n (%)	57 (49.14)	69 (57.50)	0.2473
Male, n (%)	59 (50.86)	51 (42.50)	
Age (years)			
< 65, n (%)	87 (75.00)	79 (65.83)	0.1232
≥ 65, n (%)	29 (25.00)	41 (34.17)	
Hypertension	36 (31.03)	30 (25.00)	0.3018
Diabetes	14 (12.07)	13 (10.83)	0.7656
Insomnia	16 (13.79)	29 (24.17)	0.0426
Conjunctivitis	6 (5.17)	7 (5.83)	1.0000*
Signs and symptoms			
Fever	64 (55.17)	61 (50.83)	0.5911
Fatigue	40 (34.48)	27 (22.50)	0.0579
Dry cough	70 (60.34)	64 (53.33)	0.3393
Myalgia	2 (1.72)	3 (2.50)	1.0000*
Dyspnoea	9 (7.76)	4 (3.33)	0.2285
Expectoration	13 (11.21)	11 (9.17)	0.7619
Sore throat	9 (7.76)	17 (14.17)	0.1726
Diarrhoea	22 (18.97)	15 (12.50)	0.2354
Dizziness	1 (0.86)	5 (4.17)	0.2306
Nucleic acid tests			
Positive	54 (46.55)	46 (38.33)	0.4202
Suspected	6 (5.17)	6 (5.00)	
CT (N = 235 with data)	N = 116	N = 119	0.7635
COVID-19 pneumonia	112 (96.55)	114 (95.80)	

*t test was performed for continuous variables, frequency or composition (%) were used for statistical description of classification indexes, and Chi-square test or Fisher's exact test was used for comparison between groups.

Numbers analysed

16 For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups

Total 236 patients with COVID-19 were enrolled in the full analysis set (FAS), 116 in the experimental group (favipiravir) and 120 in the control group (arbidol).

Outcomes and estimation

17a For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)

17b For binary outcomes, presentation of both absolute and relative effect sizes is recommended

The clinical recovery rate was 51.67% (62/120) in the arbidol group and 61.21% (71/116) in the favipiravir group after a 7 day's antiviral treatment (P = 0.1396), with the difference of recovery rate between two groups (95% CI) was 0.0954 (-0.0305, 0.2213).

Table 2. Comparison of 7 day's clinical recovery rate of favipiravir and arbidol in COVID-19 patients.

Variables	Favipiravir group	Arbidol group	Rate ratio (95% CI)	P value
Total patients	**(N = 116)**	**(N = 120)**		0.1396
Recovered, n (%)	71 (61.21)	62 (51.67)	0.0954 (-0.0305, 0.2213)	
Ordinary patients	**(N = 98)**	**(N = 111)**		
Recovered, n (%)	70 (71.43)	62 (55.86)	0.1557 (0.0271, 0.2843)	0.0199
Critical patients	**(N = 18)**	**(N = 9)**		
Recovered, n (%)	1 (5.56)	0 (0.00)	0.0556 (-0.0503, 0.1614)	0.4712
Patients with hypertension and/or diabetes	**(N = 42)**	**(N = 35)**		
Recovered, n (%)	23 (54.76)	18 (51.43)	0.0333 (-0.1904, 0.2571)	0.7704

Table 3 displayed duration of fever, cough relief time and auxiliary oxygen therapy or noninvasive mechanical ventilation rate between the favipiravir and arbidol groups. Of 98 ordinary COVID-19 patients in the favipiravir group, 57 had a fever and 60 had a cough; of 111 ordinary COVID-19 patients in the arbidol group, 65 had a fever and 64 had a cough. For ordinary COVID-19 patients, the time of fever reduction and cough relief in the favipiravir group was significantly shorter than that in the arbidol group (P < 0.0001). Of 42 COVID-19 patients with hypertension and/or diabetes in the favipiravir group, 28 had a fever and 25 had a cough; of 35 COVID-19 patients with hypertension and/or diabetes in the arbidol group, 24 had a fever and 23 had a cough. For COVID-19 patients with hypertension and/or diabetes, the time of fever reduction and cough relief in the favipiravir group was also significantly shorter than that in the arbidol group (P < 0.0001).

Table 3. Comparison of duration of fever, cough relief time and other secondary outcomes between two groups.

Variables	Duration of fever		Cough relief time	
	Favipiravir group	Arbidol group	Favipiravir group	Arbidol group
Ordinary patients	**N = 57**	**N = 65**	**N = 60**	**N = 64**
Day 1	12 (21.05)	2 (3.08)	1 (1.67)	3 (4.69)
Day 2	23 (40.35)	8 (12.31)	1 (1.67)	1 (1.56)
Day 3	16 (28.07)	16 (24.62)	21 (35.00)	7 (10.94)
Day 4	4 (7.02)	15 (23.08)	18 (30.00)	11 (17.19)
Day 5	0 (0.00)	13 (20.00)	9 (15.00)	12 (18.75)
Day 6	0 (0.00)	4 (6.15)	7 (11.67)	10 (15.63)
Day 7	0 (0.00)	2 (3.08)	2 (3.33)	3 (4.69)
Day 8	-	-	1 (1.67)	4 (6.25)
Day 9	-	-	0 (0.00)	1 (1.56)
Censored	2 (3.51)	5 (7.69)	0 (0.00)	12 (18.75)
Log-rank P value	< 0.0001		< 0.0001	
Patients with hypertension	**N = 28**	**N = 24**	**N = 25**	**N = 23**
Day 1	7 (25.00)	0 (0.00)	1 (4.00)	2 (9.09)
Day 2	13 (46.43)	4 (16.67)	0 (0.00)	0 (0.00)
Day 3	5 (17.86)	5 (20.83)	6 (24.00)	3 (13.64)
Day 4	3 (10.71)	2 (8.33)	7 (28.00)	2 (9.09)
Day 5	0 (0.00)	7 (29.17)	2 (8.00)	2 (9.09)
Day 6	0 (0.00)	3 (12.50)	5 (20.00)	3 (13.64)
Day 7	-	-	1 (4.00)	0 (0.00)
Day 8	-	-	2 (8.00)	2 (9.09)
Day 9	-	-	0 (0.00)	1 (4.55)
Censored	0 (0.00)	3 (12.50)	1 (4.00)	7 (31.82)

For ordinary patients with COVID-19, auxiliary oxygen therapy or noninvasive mechanical ventilation rate was 17.12% (19/111) in the arbidol group and 8.16% (8/98) in the favipiravir group (P = 0.0541), with the difference of recovery rate between 2 groups (95% CI) was -0.0895 (-0.1781, -0.0009); for critical patients with COVID-19, auxiliary oxygen therapy or noninvasive mechanical ventilation rate was 88.89 (8/9) in the arbidol group and 72.22% (13/18) in the favipiravir group (P = 0.3261), with the difference of recovery rate between 2 groups (95% CI) was -0.1667 (-0.4582, 0.1248); for COVID-19 patients with hypertension and/or diabetes, auxiliary oxygen therapy or noninvasive mechanical ventilation rate was 28.57% (10/35) in the arbidol group and 21.43% (9/42) in the favipiravir group (P = 0.4691), with the difference of recovery rate between two groups (95% CI) was -0.0714 (-0.2658, 0.1230). There was no statistical difference was observed of auxiliary oxygen therapy or noninvasive mechanical ventilation rate between 2 groups (both P > 0.05).

Of all cases enrolled in this study, the all cause mortality was 0. The rate of new dyspnea in arbidol group was 11.67% (14/120) and in favipiravir group was 3.45% (4/116) with the P value = 0.0174. The cases of respiratory failure in the two group were both 4.

Other secondary outcomes				
AOT or NMV*	**Favipiravir group**	**Arbidol group**	**Rate ratio (95% CI)**	**P value**
Ordinary patients	N = 98	N = 111		
With auxiliary, n (%)	8 (8.16)	19 (17.12)	-0.0895 (-0.1781, -0.0009)	0.0541
Patients with hypertension and/or	N = 42	N = 35		
With auxiliary, n (%)	9 (21.43)	10 (28.57)	-0.0714 (-0.2658, 0.1230)	0.4691
All-cause mortality	0 (0.00)	0 (0.00)	/	/
Dyspnea after taking	4 (3.45)	14 (11.67)	/	0.0174
Respiratory failure, n (%)	1 (0.86)	4 (3.33)	/	0.3700*

*Fisher's exact test was used for comparison between groups.

Ancillary analyses

18 Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory

Concretely, for ordinary patients with COVID-19, 7 day's clinical recovery rate was 55.86% (62/111) in the arbidol group and 71.43% (70/98) in the favipiravir group (P = 0.0199), with the difference of recovery rate between two groups (95% CI) was 0.1557 (0.0271, 0.2843); for critical patients with COVID-19, clinical recovery rate was 0 (0/9) in the arbidol group and 5.56% (1/18) in the favipiravir group (P = 0.4712), with the difference of recovery rate between two groups (95% CI) was 0.0556 (-0.0503, 0.1614); for COVID-19 patients with hypertension and/or diabetes, clinical recovery rate was 51.43% (18/35) in the arbidol group and 54.76% (23/42) in the favipiravir group (P = 0.7704), with the difference of recovery rate between two groups (95% CI) was 0.0333 (-0.1904, 0.2571) (Table 2).

Harms

19 All important harms or unintended effects in each group (for specific guidance see CONSORT for harms42)

In the whole process of trial, we detected some antiviral-associated adverse effects. 37 adverse effects cases in the favipiravir group and 28 cases in the arbidol group were observed. The most common adverse events were raised serum uric acid (3 [2.50 %] vs 16 [13.79%], P = 0.0014), more common in patients of the favipiravir group than those in the arbidol group. But no statistical difference was observed for abnormal LFT (ALT and/or AST were elevated) (12 [10.00%] in the arbidol group vs 9 [7.76%] in the favipiravir group, P =0.5455), psychiatric symptom reactions (1 [0.83%] vs 2 [1.72%]; P = 0.6171) and digestive tract reactions (nausea, anti-acid, flatulence [10]) (14 [11.67%] vs 16 [13.79%]; P = 0.6239) (Table 4). These adverse reactions disappeared when most patients were discharged from hospital.

Table 4. Comparison of antiviral-associated adverse effects between two groups.

Adverse effects	Favipiravir group (N = 116)		Arbidol group (N = 120)		P value
	Frequency	Cases, n (%)	Frequency	Cases, n (%)	
Total	43	37 (31.90)	33	28 (23.33)	0.1410
LFT abnormal	9	9 (7.76)	12	12 (10.00)	0.5455
Raised serum uric acid	16	16 (13.79)	3	3 (2.50)	0.0014
Psychiatric symptom reactions	2	2 (1.72)	1	1 (0.83)	0.6171*
Digestive tract reactions	16	16 (13.79)	17	14 (11.67)	0.6239

*Fisher's exact test was used for comparison between groups.

Discussion

Limitations

20 Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses

Inevitably, our study has some limitations. First, it was difficulty to select the drug of control group. For the COVID-19 pneumonia, there is no effective antiviral drug was reported. Chinese doctors had recommended antiviral drugs in the sixth edition of the guidelines: recombinant human interferon alfa-2b, ribavirin, chloroquine phosphate, lopinavir and arbidol. The clinical studies were currently undergoing to test the efficacy and safety of these drugs in the treatment of COVID-19. Despite the antiviral effect of arbidol, there is no exact data in the literature to support its effectiveness. Arbidol was widely used by Chinese doctors in the initial stage of antiviral epidemic of COVID-19 (Jan. 1 to Jan. 30, 2020) [11]. For ethical reasons, we chose arbidol as the positive control, and adopted the optimal experimental design. Second, due to the limitation of the observation period, it lacked the safety and effectiveness judgment as long as 1 month. Besides, it also lacked the evidence tracking of relapse (including nucleic acid conversion to positive, fever and cough again) in the next month in the discharged patients with negative nucleic acid test and normal CT imaging lung test. Third, in the inclusion criteria, we did not include the positive nucleic acid test. The accuracy of nucleic acid kit and throat swab sampling would affect the judgment of the results. We collected the number of nucleic acid positive cases in the screening period, 54 (46.55%) in favipiravir group and 46 (38.33%) in arbidol group. The clinical diagnosis and CT results suggested that there might be negative nucleic acid in patients with COVID-19 pneumonia. In the screening period, the patients with contact history, typical CT imaging results of COVID-19 and obvious clinical symptoms had negative nucleic acid test, which was related to the previous treatment, onset time, sampling and detection kit. Fourth, among all the participants, there were 18 critical patients in the favipiravir group and 9 critical patients in the arbidol group. Because of the imbalance of the proportion of critical patients between the two groups, it had an important impact on the primary

outcome (7 day's clinical recovery rate), secondary outcomes and combined medication. According to the severity of COVID-19 and whether it is combined with hypertension and/or diabetes, a stratified analysis was conducted.

Generalisability

21 Generalisability (external validity, applicability) of the trial findings

No discussion of the external validity of the trial results. The authors suggest universal applicability:

Interpretation

22 Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence

In ordinary COVID-19 patients untreated with antiviral previously, favipiravir can be considered as a preferred treatment because of its higher 7 day's clinical recovery rate and more effectively reduced incidence of fever, cough except some antiviral-associated adverse effects.

Other information

Registration

23 Registration number and name of trial registry

This study is registered with Chictr.org.cn, number ChiCTR200030254.

ChiCTR2000030254 **Note: the registry number given in the paper is missing a zero.**

Protocol

24 Where the full trial protocol can be accessed, if available

The Ethics Committee at Zhongnan Hospital of Wuhan University approved the trial protocol (approval number: 2020040) and written informed consent was obtained from all participants or their authorized representatives.

Protocol available in the supplemental information of the medRxiv submission.
https://www.medrxiv.org/content/10.1101/2020.03.17.20037432v2.supplementary-material

Funding

25 Sources of funding and other support (such as supply of drugs), role of funders

This work was supported by the National Key Research and Development Program of China (2020YFC0844400).