

# Decision Criteria During Joint Modelling of Efficacy and Safety With MCP-Mod

Caroline Kellner<sup>1</sup>, Lieven Nils Kennes<sup>1</sup>

<sup>1</sup> Department of Economics and Business Administration, University of Applied Sciences Stralsund, Stralsund, Germany

Correspondence: Lieven Nils Kennes, Department of Economics and Business Administration, University of Applied Sciences Stralsund, 18435 Stralsund, Germany. E-mail: Lieven.Kennes@hochschule-stralsund.de

Received: August 12, 2020 Accepted: December 14, 2020 Online Published: December 20, 2020

doi:10.5539/ijsp.v10n1p79

URL: <https://doi.org/10.5539/ijsp.v10n1p79>

## Abstract

MCP-Mod has been established as analysis method for investigating the dose-response (DR) relationship and dose finding in clinical Phase II trials. While most work on MCP-Mod focusses on the efficacy DR relationship, in 2015, Tao, Lin, Pinheiro and Shih (2015) extended MCP-Mod to the joint modelling of efficacy and safety endpoints in “Dose Finding Method in Joint Modeling of Efficacy and Safety Endpoints in Phase II Studies”. Their proposed algorithm defines several decision criteria, which majorly impact results or even terminate the algorithm. This viewpoint investigates the robustness of two of these decision criteria. While the criterion on the relationship between the maximum safety dose and minimum effective dose is reasonable and robust, there exist some advantages applying a more generous criterion to establish proof of concept for safety. Increasing the proposed significance level in establishing proof of concept for the safety DR relationship, helps to identify non-flat safety DR relationships which ultimately improves final estimation of the optimal dose.

**Keywords:** dose finding, efficacy, joint modelling, maximum safety dose, minimum effective dose, safety

## 1. Introduction

The aim of Phase II clinical trials in pharmaceutical development is establishing maximum information about the dose-response (DR) relationship. Traditionally multiple comparison procedures (MCP) or modelling techniques (Mod) were applied to derive information on the DR relationship. In 2005, Bretz, Pinheiro and Branson (2005) introduced a novel dose-finding method, MCP-Mod, combining these two traditional approaches. The main idea is to apply the modelling approach under model uncertainty by using a pre-defined candidate set of potential functional DR relationships. Since then, MCP-Mod is gaining increasing popularity and is nowadays applied in various industry settings (Verrier, Sivapregassam, & Solente, 2014); (Mercier, Bornkamp, Ohlssen, & Wallstroem, 2015); (Kennes, Volkens, & Kralidis, 2019). A Qualification Opinion of the European Medicines Agency on MCP-Mod (European Medicines Agency (EMA); Committee for Medicinal Products for Human use (CHMP), 2014) acknowledged MCP-Mod to be an “efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty.” The US Food and Drug Administration communicates MCP-Mod as a Fit-for-Purpose statistical approach for dose-finding (Food and Drug Administration, 2016) aiming to facilitate greater utilization of this method in drug development programs.

The scientific literature on MCP-Mod mainly focusses on efficacy endpoints, establishing a dose-efficacy-relationship. However, to obtain full knowledge of the benefit-risk profile, a simultaneous functional relationship including both components, efficacy and safety, is desirable. Combining efficacy and safety in dose-finding studies is not new outside the framework of MCP-Mod, e.g. by using copula models (Tao, et al., 2013); (Deldossi, Osmetti, & Tommasi, 2016). However, while more than 300 papers address MCP-Mod, only Tao, Lin, Pinheiro and Shih (2015) extended the MCP-Mod approach to select the best joint model based on two correlated outcomes, efficacy and safety. Due to the efficiency of MCP-Mod and the advantages of simultaneously deriving information on efficacy and safety, the proposed procedure is very promising. Aim of this research is to investigate the proposed procedure in greater depth. The original algorithm of Tao, Lin, Pinheiro and Shih (2015) defines several stopping criteria, which directly impact the results or even terminate the algorithm. Two of these stopping criteria will be investigated in an extensive simulation study, in particular regarding their impact on study results.

## 2. Theoretical Background

To simultaneously obtain knowledge of the efficacy and safety profiles, Tao, Lin, Pinheiro and Shih (2015) proposed a

bivariate joint model with two different functional DR relationships for the efficacy and safety component and a potentially correlated bivariate normal error term. The functional DR relationship for efficacy and for safety is derived via MCP-Mod, either by separate model fitting or joint model fitting. At this point, for the efficacy DR relationship, the minimum effective dose (MED) and for the safety DR relationship the maximum safety dose (MSD) can be derived. The resulting bivariate joint model is then studied further to determine the optimal dose. Either the difference between the standardized versions of the efficacy and safety functions, in the following called utility function, is maximized or the joint success probability for Phase III is maximized based on success-parameters pre-defined for Phase III. The resulting dose is the final optimal target dose for the subsequent Phase III study. For details on the above steps we refer to the original paper (Tao, Lin, Pinheiro, & Shih, 2015). For the research of this manuscript, we chose to maximize the utility function to obtain the final optimal dose, thus their approach can be summarized by four steps:

- 1) MCP-Mod is performed on the efficacy endpoint at a significance level of  $\alpha_1 = 0.05$ . If proof of concept (PoC) is not established, the algorithm terminates as the efficacy DR relationship appears to be flat and there is no therapeutic potential.
- 2) If PoC is established in 1), MCP-Mod is performed on the safety endpoint at a “significance level” of  $\alpha_2 = 0.2$ .
- 3) If no PoC is established in 2), the safety profile appears to be flat and only the dose-efficacy relationship is studied further to identify the best dose for Phase III. If PoC is established in 2) and the MSD is larger than or equal to the MED, joint modeling for efficacy and safety is performed.
- 4) The optimal dose (target dose) is selected maximizing a utility function, i.e. maximizing the differences between the standardized versions of the efficacy and safety functions determined in 3).

Each step, including their stopping criteria is sensible. However, MED and MSD may vary based on the formula applied to derive them. Already in the original MCP-Mod Paper (Bretz, Pinheiro, & Branson, 2005), three different sensible formulas of MED are proposed. The stopping criteria  $\overline{MSD} < \overline{MED}$  might be affected by the choice of the formula or simply by the precision of the estimates. In a first investigation, the robustness of this stopping criteria is investigated. Second, the choice of  $\alpha_2 = 0.2$  of the above mentioned step 2) appears to some extent arbitrary and different choices of  $\alpha_2$  and their impact on final results are investigated.

### 3. Simulation Model

In an extensive simulation study, both criteria, the stopping criteria  $\overline{MSD} < \overline{MED}$  and the choice of  $\alpha_2$ , are investigated, in particular regarding their consequences on final results. Phase II clinical trials including 240 subjects were repeatedly simulated under different scenarios. Subjects were divided equally among four dose levels. For the purpose of this investigation we chose equidistant dose levels between 0 (Placebo) and 600 based on a recent real world MCP-Mod investigation (Kennes, Volkens, & Kralidis, 2019). The dose level is denoted by  $d_j \in \{0, 200, 400, 600\}$ ,  $j = 1, \dots, 4$ . Patient data for efficacy and safety was generated for each subject based on the following e-max and exponential model (resp.):

- 1.) E-max DR-relationship for the efficacy endpoint:

$$f(d, \theta_Y) = \theta_0 + \theta_1 d / (\theta_2 + d), \quad \theta_0 = 2, \theta_1 = 1.2, \theta_2 = 160 \quad (1)$$

- 2.) Exponential DR-relationship for the safety endpoint:

$$g(d, \theta_Z) = \theta_0 + \theta_1 \exp(d / \delta), \quad \theta_0 = 0, \theta_1 = 0.8 / \exp(3), \delta = 200 \quad (2)$$

The parameters  $\theta_0, \theta_1, \theta_2$  and  $\delta$  are chosen to reflect sensible outcome values of a recent real world MCP-Mod investigation in the therapeutic area of chronic pain (Kennes, Volkens, & Kralidis, 2019). The two mathematical functions in (1) and (2) constitute the true, underlying DR relationships for efficacy and safety, MCP-Mod is aiming to detect. However, to model heterogeneity in individual patient response, the above function value is modified by an additive random normal noise component. To investigate different magnitudes of noise, in scenario 1, the standard deviation of the random error component for efficacy and safety is 0.87, 0.9, respectively, while in scenario 2 it is 2.6, 2.7, respectively.

To detect the efficacy and safety DR relationship under model uncertainty, the MCP-Mod candidate set consists of four models: E-max, exponential, logistic and linear. Thus in both cases the correct functional relationship is included. The parameters of the standardized version of the DR-models of the candidate set are 160 (e-max), 204.57 (exponential) and {215.08, 52.66} (logistic) (Bretz, Pinheiro, & Branson, 2005). The best model is selected by Akaike's Information Criterion (AIC). Target doses of interest are MED and MSD, calculated by

$$MED = \operatorname{argmin}\{f(d, \theta_Y) > f(d_1, \theta_Y) + 0.7\} \quad (3)$$

and

$$MSD = \operatorname{argmin}\{g(d, \theta_z) > g(d_1, \theta_z) + 0.5\}. \tag{4}$$

The Method to determine the optimal dose is based on the utility function  $F(d) = \operatorname{eff}(d) - k * \operatorname{saf}(d)$ , where  $k = 0.9$  is a weight to discount safety for efficacy. The functions  $\operatorname{eff}(d)$  and  $\operatorname{saf}(d)$  are standardized mean responses according to Tao, Lin, Pinheiro and Shih page 38 (2015). The true optimal dose maximizing the above utility function is 354 for scenario 1 (low variance) and 355 for scenario 2 (high variance), respectively. For each scenario, 10 000 trails were simulated.

#### 4. Results

First we investigated whether a trial in our simulation study exists where  $\widehat{MSD}$  is just below  $\widehat{MED}$ , thus the algorithm in Tao, Lin, Pinheiro and Shih (2015) would terminate, however due to the specific dose-response-profiles, a certain dose level might yield only a small loss of efficacy compared to the  $\widehat{MED}$  but a large improvement in safety compared to the  $\widehat{MSD}$ . Figure 1 illustrates such a theoretical scenario.  $\widehat{MSD} < \widehat{MED}$ , but a large difference between efficacy and safety might yield a desirable benefit-risk-ratio for the optimal dose, while efficacy is obtained at a similar and only slightly lower level compared to  $\widehat{MED}$ . A slightly lower level of  $\widehat{MED}$  is accepted in this investigation due to the above described variation in formulas and the potential deviation from the true value due to unsystematic estimation errors.

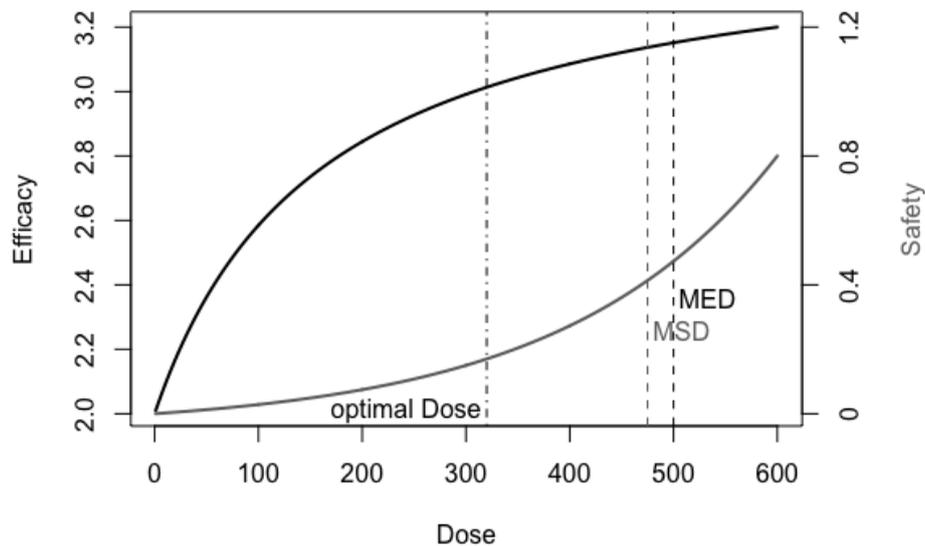


Figure 1. Dose-response profiles for efficacy and safety with  $MSD < MED$ , but potentially useful benefit-risk-ratio

For both scenarios, all trials with  $0.7 \cdot \widehat{MED} < \operatorname{optimal\ dose} < \widehat{MSD} < \widehat{MED}$  were extracted and investigated. The value 0.7 was chosen in the inequation to only obtain doses that are somewhat near the  $\widehat{MED}$ . In scenario 1, only 6 trials (0.06%) fulfilled the above criterion. For these trials, in fact the  $\widehat{MSD}$  is only slightly below  $\widehat{MED}$  (mean: -17.1, range: [-33.1, -2.8]) and the difference in efficacy between the optimal dose and  $\widehat{MED}$  is low (mean: -0.1122) while the difference in safety between the optimal dose and  $\widehat{MSD}$  is large (mean on smaller scale: -0.2926). The estimated optimal dose is 363.83 on average (range [355-371]), and thus always in close proximity to its true parameter 354. One notes that in all these trials, a linear instead of the true e-max dose-response relationship was selected for efficacy. For safety, the true exponential dose-response relationship was selected in three, a logistic model in the remaining three cases. The left graph of Figure 2 illustrates the DR-curves of one of these 6 trials as an example. For efficacy a linear model, for safety a logistic model was selected, thus in both choices being incorrect.  $\widehat{MSD}$  is only slightly below  $\widehat{MED}$ , but due to the steep slope of the fitted logistic safety function just prior to these two dose estimates, a large difference between efficacy and safety is established at a dose level of approx. 375. For this dose level efficacy is established only slightly below the response value of the  $\widehat{MED}$ .

In scenario 2, 24 trials (0.24%) fulfilled the above criterion. The difference between  $\widehat{MSD}$  and  $\widehat{MED}$  is larger on average (mean: -26.36, range: [-67.0324, -0.0028]), however the differences in efficacy (mean: -0.1115) and safety (mean: -0.2871) is comparable to scenario 1. The estimated optimal dose was  $281.58 \pm 89.55$ , and thus in scenario 2 not always in close proximity to its true parameter 355. The correct model combination was chosen in none of the 24 trials. The choices were dominated by the combination linear/logistic (58.33 %), followed by exponential/linear (25%) and lastly linear/exponential, exponential/exponential (each 8.33 %). The right graph of Figure 2 illustrates the DR-curves

of one of these 24 trials as an example. For efficacy a linear model (incorrect choice), for safety an exponential model (correct choice) was selected.  $\overline{MSD}$  is only slightly below  $\overline{MED}$  and due to the exponential course of the safety model, a large difference between efficacy and safety is established at a dose level of approx. 450. For this dose level efficacy is established only slightly below the response value of the  $\overline{MED}$ .

In summary, for a low and a high value of noise, the relative frequency of potential meaningful cases outside the stopping criterion  $\overline{MSD} < \overline{MED}$  is low. The sensible stopping criterion is found to be robust against the above described deviation.

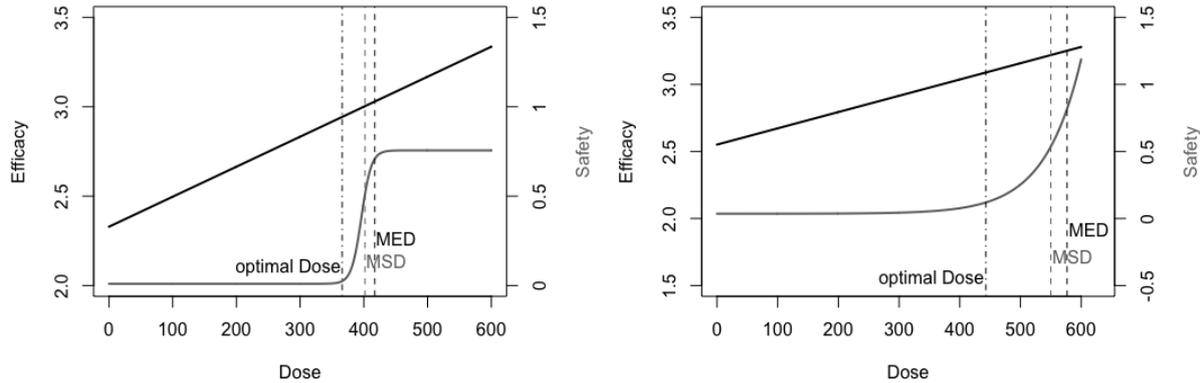


Figure 2. DR-profiles for efficacy and safety. Left: Scenario 1. Right: Scenario 2

The second investigation focused on the significance level  $\alpha_2 = 0.2$ , chosen to establish PoC for the safety-profile. The choice of  $\alpha_2$  appears to some extent arbitrary and different values of  $\alpha_2$  between 0.1 and 0.6 were investigated, to gain insight in a broad range of values for  $\alpha_2$ . According to the original proposal by Tao, Lin, Pinheiro and Shih (2015), p-values larger than 0.2 would yield an optimal dose of 600 due to the conclusion of a flat safety DR profile and the monotonicity of the efficacy DR profile. In a first step the number of trials with safety p-values larger than 0.2 were determined for both scenarios. In scenario 1, only two of the 10 000 simulated trials, yielded a p-value larger than 0.2 ( $p \in \{0.24, 0.27\}$ ). The model selection was e-max/linear in both cases and the optimal dose determined as 285 and 600. Due to this low number of occurrences, the choice of the  $\alpha_2$ -value has no noteworthy impact on results in scenario 1 in terms of overlooking non-flat DR profiles.

In scenario 2, 2932 trials yielded a p-value larger than 0.2, thus about 30% of simulated trials. As the true safety DR relationship is in fact not flat, a value of  $\alpha_2 = 0.2$  appears to be too restrictive. Table 1 compares certain performance measures for different values of  $\alpha_2$ . For example, setting  $\alpha_2 = 0.5$ , 1640 additional trials fulfilled the modified PoC-criterion (p-value lower than 0.5) and thus did not automatically yield an optimal dose of 600. However only 983 of these trials completed the whole algorithm. Additionally, 27 trials with a safety p-value below 0.2 which were previously not successful, now complete the whole algorithm, while 142 trials that completed the whole algorithm for  $\alpha_2 = 0.2$ , did not fully complete it for  $\alpha_2 = 0.5$ . Concerning the optimal dose estimation, a less strict value for  $\alpha_2$  heavily improves the average estimate yielding at least on average a very small difference to its true parameter value 355. The number of correct model selection for both models (efficacy and safety) is maximized at  $\alpha_2 = 0.5$  (Table 1).

Table 1. Performance of the algorithm to determine the optimal dose for different values of  $\alpha_2$  in Scenario 2

$\alpha_2$	0.1	0.2	0.3	0.4	0.5	0.6
Number of trials with fully completed algorithm	2142	2716	3076	3306	3584	3860
Number of trials with incompleted algorithm	7858	7284	6924	6694	6416	6140
- $\overline{MSD} < \overline{MED}$	1032	1124	1160	1166	1140	1130
- $\overline{MSD}$ not available	9	26	59	116	203	292
- optimal dose outside $[\overline{MED}, \overline{MSD}]$	564	932	1223	1439	1511	1500
- $\text{effpVal} > 0.05 (= \alpha_1)$ or $\text{MED} > 600$	2270	2270	2270	2270	2270	2270

- safpVal> $\alpha_2$	3983	2932	2212	1703	1292	948
Correct selection of both models (e-max/exp)	868	885	891	900	911	905
Correct selection of efficacy model (e-max)	3942	3942	3942	3942	3942	3942
Correct selection of safety model (exp)	2205	2257	2294	2315	2322	2321
Number of trials with optimal dose 600	3983	2932	2212	1707	1405	1213
Optimal dose estimate (average)	482.77	439.4	404.63	380.69	362.69	351.15
Optimal dose estimate (standard deviation)	189.13	205.2	211.63	211.27	211.27	209.19

## 5. Discussion

Two decision criteria of the algorithm proposed by Tao, Lin, Pinheiro and Shih (2015) were investigated. If the first criterion,  $\widehat{MED} < \widehat{MSD}$  is not fulfilled, only few cases might yield a to some extent useful therapeutic dose-response profile despite  $\widehat{MSD} < \widehat{MED}$  due to specific functional relationships. However, in none of these cases both models (efficacy and safety) were chosen correctly, indicating deviance from the actual relationship. From our experience, in a specific, real world trial, the obtained single result is anyhow investigated closely and such a finding would be discussed explicitly. Especially due to the low number of occurrences, this first investigated stopping criterion appears to be robust. Regarding the significance level to establish PoC of the safety profile, our simulation study shows, that choosing a less strict value for  $\alpha_2$  performs superior on average, especially for data with higher variation. A higher percentage of correct model selections is observed and more trials complete the algorithm. The actual non-flat safety profile is detected more often and thus those trials did not automatically receive the upper dose range value as optimal dose due to the monotonicity of the efficacy DR relationship. These intermediate findings ultimately result in a more accurate estimation of the final optimal dose, while unfortunately a large standard deviation of the estimates persist. Altogether a higher value for  $\alpha_2$ , e.g.  $\alpha_2 = 0.5$ , is advocated. Such a choice will more likely prevent overlooking an actual increase in toxicity and is in line with the usual conservative approach in pharmaceutical development.

## 6. Conclusion

Tao, Lin, Pinheiro and Shih (2015) developed a meaningful extension of the MCP-Mod methodology to simultaneously derive information on the DR relationship regarding efficacy and safety. Information on both DR relationships enable better decision making for selecting the target dose in subsequent Phase III trials. Their proposed procedure appears to be robust regarding the stopping criterion,  $\widehat{MSD} < \widehat{MED}$ , however would profit to some extent from choosing a larger value of  $\alpha_2$  to prevent overlooking an actual increase in toxicity and ultimately improve final estimation of the optimal dose.

## References

- Bretz, F., Pinheiro, J. C., & Branson, M. (2005). Combining Multiple Comparisons and Modeling Techniques in Dose-Response Studies. *Biometrics*, (61), 738-748. <https://doi.org/10.1111/j.1541-0420.2005.00344.x>
- Deldossi, L., Osmetti, S. A., & Tommasi, C. (2016). PKL-optimality criterion in copula models for efficacy-toxicity response. In *mODa 11-Advances in Model-Oriented Design and Analysis* (pp. 79-86). Springer. [https://doi.org/10.1007/978-3-319-31266-8\\_10](https://doi.org/10.1007/978-3-319-31266-8_10)
- European Medicines Agency (EMA); Committee for Medicinal Products for Human use (CHMP). (2014). *Qualification opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty*. Retrieved from [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Regulatory\\_and\\_procedural\\_guideline/2014/02/WC500161027.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2014/02/WC500161027.pdf)
- Food and Drug Administration. (2016). *Drug development tools: fit-for-purpose initiative*. Retrieved from <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/ucm505485.htm>
- Kennes, L. N., Volkens, G., & Kralidis, G. (2019). Study design aspects and inter-subject variability in longitudinal clinical phase II dose-finding trials. *Pharmaceutical statistics*, 18(2), 248-259. <https://doi.org/10.1002/pst.1921>
- Mercier, F., Bornkamp, B., Ohlssen, D., & Wallstroem, E. (2015). Characterization of dose-response for count data using a generalized MCP-Mod approach in an adaptive dose-ranging trial. *Pharmaceutica Statistics*, 14(4), 359-367. <https://doi.org/10.1002/pst.1693>
- Tao, A., Lin, Y., Pinheiro, J., & Shih, W. J. (2015). Dose Finding Method in Joint Modeling of Efficacy and Safety Endpoints in Phase II Studies. *International Journal of Statistics and Probability*, 4(1), 33-45. <https://doi.org/10.5539/ijsp.v4n1p33>

- Tao, Y., Liu, J., Li, Z., Lin, J., Lu, T., & Yan, F. (2013). Dose-finding based on bivariate efficacy-toxicity outcome using archimedean copula. *PloS one*, *11*(8), e78805. <https://doi.org/10.1371/journal.pone.0078805>
- Verrier, D., Sivapregassam, S., & Solente, A. C. (2014). Dose-finding studies, MCP-Mod, model selection, and model averaging: two applications in the real world. *Clinical Trials*, *11*(4), 476-484. <https://doi.org/10.1177/1740774514532723>

### **Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).