

## ORIGINAL ARTICLES

# Statistical modeling and verification for the synthesis of median survival time in multilevel meta-analysis of survival data

Jiajie Zang<sup>1,2</sup>, Jinfang Xu<sup>2</sup>, Chun Xiang<sup>2</sup>, Shurong Zou<sup>1</sup>, Jia He <sup>\*2</sup>

<sup>1</sup>Department of Nutrition Hygiene, Institute of Health Related Risk Factor Identification and Control, Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China

<sup>2</sup>Department of Health Statistics, Second Military Medical University, Shanghai, China

**Received:** June 5, 2015

**Accepted:** July 9, 2015

**Online Published:** July 22, 2015

**DOI:** 10.5430/jer.v1n1p25

**URL:** <http://dx.doi.org/10.5430/jer.v1n1p25>

## ABSTRACT

**Objective:** Past meta-analyses of survival data have been over simplistic because of restricting to proportional hazard model, lack of intuitive results, and potential omitting information. These had the potential to recommend sub-optimal policies. Here we develop multilevel methods for combining median survival times (MSTs) for meta-analysis of survival data.

**Methods:** We used simulated data to test and verify the synthesis model we developed. We generated the study-level data to fit multilevel model and individual patient data to calculate gold standard. We then used the Bland-Altman method and the relative change from the gold standard to evaluate the fit of the models. Examples were presented in a meta-analysis to illustrate the feasibility of the models.

**Results:** We generated eight sets of simulated datasets of different number of studies and sample size. We established the multi-level fixed and random effect models to pool the MSTs. The test of the fitness of the model showed that the means of difference (d) for all simulated datasets between the calculated values and the gold standards are no more than -0.230 and -0.329 days and the largest 95% CIs of d are -3.823 ~ 3.364 and -3.936 ~ 3.278 days respectively. At least 91.9% and 92.3% of the difference between the estimated values and the gold standards are small. The real examples of a meta-analysis were provided with combined MSTs along with pooled HR.

**Conclusions:** The multilevel models of synthesizing MSTs in survival data AD meta-analysis were verified with good fitting effects and provide more intuitive information.

**Key Words:** Median survival time, Pooled, Survival data, Multilevel model

## 1. INTRODUCTION

In evidence-based medical research, individual patient data (IPD) meta-analyses, which involve central collection, checking, and reanalysis of updated IPD, have been described as the benchmark of systematic review.<sup>[1,2]</sup> Although IPD

meta-analyses are increasing,<sup>[3]</sup> but practitioners may face the practical problem of obtaining IPD from only a portion of the studies<sup>[4]</sup> due to economic, resource, or time constraints. When IPD may not be available, and only aggregated data (AD), such as a treatment effect estimate and its standard

\*Correspondence: Jia He; Email: [jjzang@scdc.sh.cn](mailto:jjzang@scdc.sh.cn); Address: Department of Health Statistics, Second Military Medical University, Shanghai, China.

error, can be obtained from some studies.<sup>[5]</sup> These AD are mainly extracted from the published literature.

The most commonly used method in meta-analysis is the synthesis of AD obtained from study publications or study authors. For time-to-event outcomes such AD often include the hazard ratio (HR) and its variance, which are calculated to present the overall instantaneous risk of an event in an experimental intervention compared with a control.<sup>[6]</sup> Meta-analysis produces a weighted average of the AD across studies to give an overall measure of treatment effect in the fixed effect model.<sup>[7]</sup> It is also combined in a random effect meta-analysis that accounts for between-study heterogeneity in treatment effect.<sup>[8]</sup>

The median survival times (MSTs) and the combined HR and its 95% confidence interval (CI) are available from the IPD meta-analysis of time-to-event data,<sup>[3,9]</sup> which could become a thorough survival analysis by assembling every individual patient from each trial. However, AD meta-analyses usually only calculate the combined HR and its 95% CI without pooled MSTs.

In these circumstances AD meta-analysis in evidence-based medical research still dominates in the meta-analysis of survival data. There are some problems when conducting AD meta-analysis. Firstly, the synthesis of AD should be under the hypothesis of proportional hazards regression models for original data. However, not all the survival analysis can meet such a hypothesis, which may introduce bias for the pooled results.<sup>[10]</sup> Secondly, the combined HRs only conveys the relative hazard risk of the event between the experiment and the control groups. Researchers or individuals must evaluate the efficacy of treatment from an intuitive index. Clinicians often are more concerned about the direct effect of a treatment rather than the relative hazard risk. Thirdly, the survival curves and MSTs might be different in each group when the same HRs and 95% CIs are combined from the different meta-analysis of AD. Due to the drawbacks of pooled HRs in survival data meta-analysis, MSTs can provide more direct information of the effect of a treatment. However, an effective method of combining MSTs for each group, which could be gathered from an IPD meta-analysis of survival data, has not yet been developed for AD meta-analysis.

Multilevel meta-analysis has been used to adjust for nonequivalent study results in meta-analytic research and to gain substantive information for evaluating treatment efficacy.<sup>[11]</sup> It also used to statistically adjust for nonequivalent or nonrandomized groups in observational research and adjust covariates or the study.

The purpose of this study is to develop alternative multilevel

methods to combine MSTs for meta-analyses of survival data, which may be more readily available in publications and a more direct index in survival analysis.

## 2. SOURCE OF DATA

### 2.1 Simulated data

We simulated the individual patient time-to-event (survival) data following exponential distribution.<sup>[12]</sup> The simulated variables included patient number, group, survival time, and survival outcome. We manipulated the raw data in two ways. On the one hand, we calculated study-level AD from the simulated patient data. We used a proportional hazard regression model (Cox model) for constant hazards to analyze HRs and their 95% CIs of experimental effects. We used a linear interpolation method to estimate the MSTs. These AD represent the data extracted from published literature when conducting a real meta-analysis. We used the summarized data to fit the model for the synthesis of MSTs. On the other hand, we combined the raw data to construct the IPD dataset. We gathered and estimated the simulated data from each study included in a meta-analysis to obtain the IPD HR, its 95% CI, and the MST. We considered the IPD MSTs of each group to be the gold standard and compared them with the MSTs computed from the model we developed in this study. The detailed information was shown in our previous study.<sup>[13,14]</sup>

We named the eight simulated datasets of meta-analyses we generated M20S750, M20S500, M10S500, M10S300, M10S200, M5S400, M5S300, and M5S200. M refers to meta-analysis; the following number refers to the number of studies included in the meta-analysis; S refers to study; and the following number refers to the number of individual patients included in the study. Each dataset includes 15,000, 10,000, 5,000, 3,000, 2,000, 1,500, or 1,000 individual patients, and each meta-analysis includes 20, 10, or 5 studies. We derived the IPD and trial-level AD statistics from the simulated data.

We calculated the MSTs from the IPD datasets as the gold standard and used the trial-level MSTs to fit the model constructed later. We conducted the simulations 500 times to obtain the gold standard and to compute MSTs from the models. Table 1 shows the simulation datasets.

### 2.2 Example data: Meta-analysis of survival data

We conducted the real meta-analysis, including the synthesis of MSTs, using multilevel data extracted from literature about the efficacy of small-molecule epidermal growth factor receptor, tyrosine kinase inhibitor (EGFR TKIs) on non-small-cell lung cancer (NSCLC) patients in different treatment programs. We collected those studies from the general databases Pubmed and Embase and from the trial registration

websites. We used the Parmar method to pool the HRs and their 95% CIs<sup>[9]</sup> and to combine the MST and the median progression time.

### 2.3 Model evaluation

Using the Bland-Altman method and the relative changes of the model value from the gold standard, we compared the consistency between the MSTs derived from our model and the gold standard values of IPD meta-analyses.

**Table 1.** Details of the simulated datasets

Dataset	Number of studies in each meta-analysis	Number of patients in each study	Number of patients		Total number of patients in each meta-analysis
			Experimental group	Control group	
M20S750	20	750	500	250	15,000
M20S500	20	500	250	250	10,000
M10S500	10	500	250	250	5,000
M10S300	10	300	100	200	3,000
M10S200	10	200	100	100	2,000
M5S400	5	400	200	200	2,000
M5S300	5	300	150	150	1,500
M5S200	5	200	100	100	1,000

## 3. MULTILEVEL MODEL META-ANALYSIS OF MST

### 3.1 Normalization of the effect size

We transferred the MST data into a logarithm to construct the model of the meta-analysis. In this study the MSTs are combined from the individual studies. In a meta-analysis the effect sizes are assumed to be approximately normally distributed.<sup>[6]</sup> For this study we transformed the MSTs of both the experimental and the control groups to the logarithm MSTs ( $\log MST$ ) to approach a normal distribution due to its exponential distribution.<sup>[15]</sup> The effect size of the MSTs of the experimental and the control groups are defined as  $Y_1$  and  $Y_2$ , respectively. It holds that

$$Y_1 = \log(Median_1), Y_2 = \log(Median_2) \quad (1)$$

where  $Median_1$  and  $Median_2$  denote the MST of the experimental group and the control group, respectively.

### 3.2 Fitting the multilevel model meta-analysis of MST

We formulated a general class of meta-analysis models to MST synthesis by considering a simple two-level structure. It is assumed that we have a collection of studies, each concerned with the comparison of two treatment groups. The basic models we developed are “variance component” models with the variance heterogeneity cases incorporated.<sup>[16]</sup>

For the  $i$ th subject in the  $j$ th study who received the treatment, we can write a basic underlying model for outcome  $y_{ij}$  of IPD as

$$y_{ij} = \beta_0 + \mu_{0j} + e_{ij}; \mu_{0j} \sim N(0, \sigma_\mu^2); e_{ij} \sim N(0, \sigma_e^2) \quad (2)$$

Where  $j = 1, 2, \dots, k$  refers to the level 2 unit, and where  $i = 1, 2, \dots, n_j$  refers to the level 1 unit.  $\beta_0$  is the average effect of interest,  $\mu_{0j}$  is the random effect of the treatment for the  $j$ th study, and  $e_{ij}$  is the random residual of the treatment for subject  $i$  in study  $j$ .

When using the AD from the literature, the model can be written as

$$y_{.j} = \beta_0 + \mu_{.j} + e_{.j} \quad (3)$$

where  $y_{.j} = \sum_{i=1}^{n_j} y_{ij}/n_j$ ,  $\beta_0$  is the average effect of interest,  $u_{.j} = N(0, \sigma_u^2)$  represents the between-study heterogeneity,  $e_{.j} = N(0, \sigma_e^2/n_j)$  is the within-trial sampling error.  $Var(\mu_{.j}) = \sigma_u^2 = \tau^2, Var(e_{.j}) = var(\sum e_{ij}/n_j) = \sigma_e^2/n_j$  (to constrain  $\sigma_e^2 = 1$ ).

### 3.3 Study weight when fitting the models

#### 3.3.1 Study weight for model 1

We deemed the  $\log MST$  as a continuous measurement of each treatment group. We used SAS statistical software to fit the multilevel model of the MST. In SAS the model is available in a PROC MIXED model for continuous outcome measures, which fits the structure we selected to the data with restricted maximum likelihood (REML) (Leyland and Groenewegen, 2003). We set up one kind of weight for the experimental group as  $\sqrt{n_{i1}}$  and for the control group as  $\sqrt{n_{i2}}$  because they are commonly used and defined the

multilevel model weighted in this method as model 1.

### 3.3.2 Study weight for model 2

We set up another kind of weight for the experimental group as  $SE_1 = \sqrt{\frac{n_2}{n_1+n_2}} \times \frac{\log(HRU)-\log(HRL)}{3.92}$  and for the control group as  $SE_2 = \sqrt{\frac{n_1}{n_1+n_2}} \times \frac{\log(HRU)-\log(HRL)}{3.92}$  because the variance of  $\log MST$  results not only from the sample size of the study but also from the study's precision.  $\frac{n_2}{n_1+n_2}$  and  $\frac{n_1}{n_1+n_2}$  are the weight components of the sample size in each group of a study, and  $\frac{\log(HRU)-\log(HRL)}{3.92}$  represents the study precision of a survival analysis. We used these two weighted methods to statistically fit the multilevel models of MSTs and defined the multilevel model weighted in this method as model 2 (HRU, hazard ratio upper confidence interval; HRL, hazard ratio lower confidence interval).

## 4. MODEL EVALUATION RESULTS

### 4.1 Bland-Altman method

Using the Bland-Altman method, we compared the calculated MSTs derived from the model with the gold standard MSTs from the IPD datasets. For the multilevel model 1 of the M20S750 dataset, the mean of difference ( $d$ ) between the calculated value and the gold standard is -0.052, and the 95% CI of ( $d$ ) is from -1.087 to 0.984. This indicates that 95% of the derived value from our model is no more than 0.984 or less than -1.087 of the gold standard value. In other words, there is good agreement between our model method and the gold standard method (the IPD results). For the multilevel model 2, the mean of ( $d$ ) between the calculated value and the gold standard is -0.105, and the 95% CI of ( $d$ ) is from -1.154 to 0.944. This indicates good agreement between the calculated value using a random effect model and the gold standard method.

Table 2 shows the Bland-Altman results in measurement agreements for the other datasets. All of the results indicate good agreement between estimated values with both fixed and random effect models and the gold standard values for the IPD datasets.

### 4.2 Relative change method

We also calculated the relative changes ( $p_d$ ) of the model value from the gold standard ( $g$ ) for each dataset.

$$p_d = d/g \times 100\% \tag{4}$$

$d$  represents the difference between the calculated value and

the gold standard. One hundred percent of the  $d$  value is within 5% of the gold standard value for the dataset M20S750 when using multilevel models 1 and 2, which means the variation between the calculated value and the gold standard value is no more than 5% of the gold standard. Table 3 presents the details.

**Table 2.** Results of the Bland-Altman method comparing the agreement between values derived from the multilevel model and the gold standard from IPD

Dataset	Model 1		Model 2	
	Md	95% CI of $d$	Md	95% CI of $d$
M5S200	-0.230	-3.823~3.364	-0.329	-3.936~3.278
M5S300	-0.083	-3.311~3.145	-0.187	-3.424~3.050
M5S400	-0.144	-2.433~2.146	-0.194	-2.479~2.090
M10S200	-0.178	-2.698~2.342	-0.313	-2.845~2.220
M10S300	-0.097	-2.907~2.713	-0.152	-2.977~2.674
M10S500	-0.073	-2.649~2.503	-0.133	-2.710~2.444
M20S500	-0.084	-1.181~1.013	-0.140	-1.245~0.965
M20S750	-0.052	-1.087~0.984	-0.105	-1.154~0.944

Note. Multilevel model 1 weighted by  $\sqrt{n_{i1}}$  and  $\sqrt{n_{i2}}$  for experimental and control groups,

respectively. Multilevel model 2 weighted by  $SE_{n1} = \sqrt{\frac{n_2}{n_1+n_2}} \times \frac{\log(HRU)-\log(HRL)}{3.92}$  and

$SE_{n2} = \sqrt{\frac{n_1}{n_1+n_2}} \times \frac{\log(HRU)-\log(HRL)}{3.92}$  for experimental and control groups, respectively.

Md: mean of the difference between calculated median survival times and gold standard

### 4.3 Example of a meta-analysis

For our actual meta-analysis of survival data we retrieved data from trials that compared the most frequently used small-molecule EGFR TKIs for maintenance treatment of patients with NSCLC with control groups. The effects of two maintenance treatment we compared were continuous maintenance and switch maintenance. We gathered data on overall survival (OS), progression-free survival (PFS), the MST, and the median progression time from the trials and summarized them. We identified nine trials<sup>[17-26]</sup> that included 6,655 patients.

When we conducted the meta-analysis in the traditional way, the results showed that continuous maintenance with small-molecule EGFR TKIs failed to improve OS (HR 1.05; 95% CI 0.98-1.14;  $p = .186$ ) or PFS (HR 0.88; 95% CI 0.74-1.05;  $p = .165$ ). In addition we combined the MSTs using the method we developed and described above and found 10.24 months for the experimental group and 10.55 months for the control group. We found the median progression time to be 5.44 months for the experimental group and 5.37 months for the control group (see Figure 1).

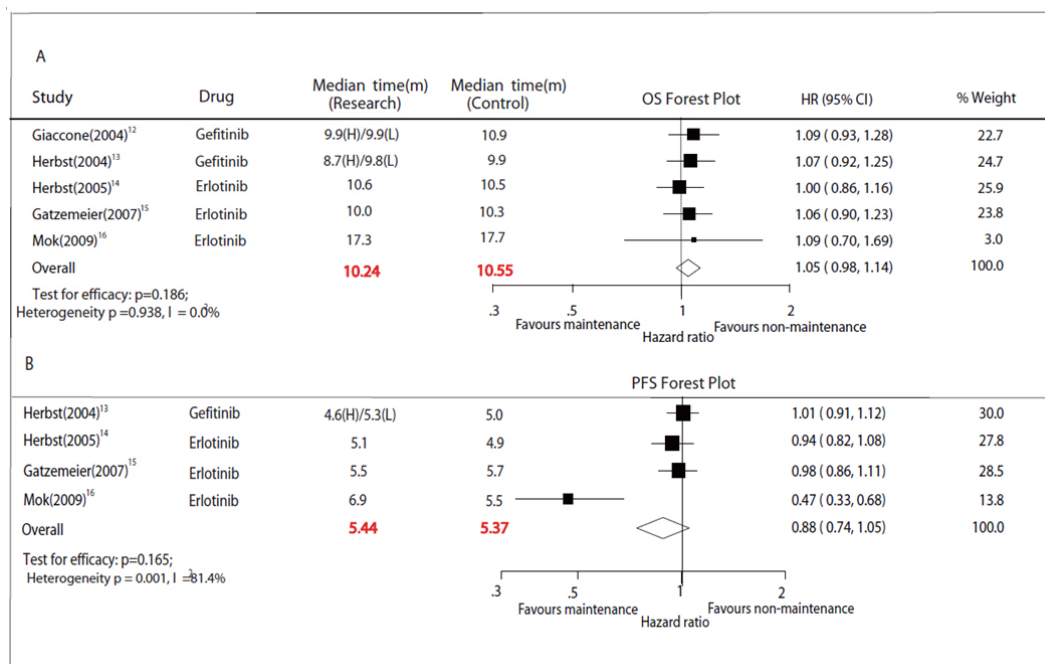


Figure 1. Combined MSTs and HRs of OS (A) and PFS (B) of continuous maintenance therapy and placebo/observation

In contrast, switch maintenance with small-molecule EGFR TKIs significantly improved OS (HR 0.87; 95% CI 0.80-0.95;  $p = .001$ ) and PFS (HR 0.75; 95% CI 0.69-0.81;  $p < .001$ ) when we combined the HRs and their variances. Ad-

ditionally, the MST was 12.86 months for the experimental group and 11.44 months for the control group and the median progression time 3.53 months for the experimental group and 2.67 months for the control group (see Figure 2).

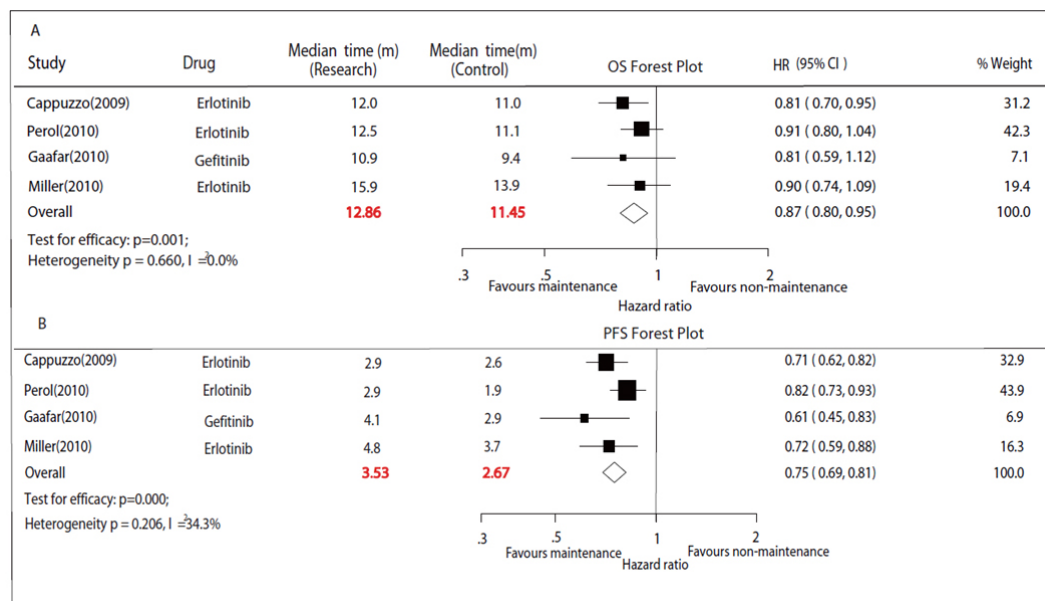


Figure 2. Combined MSTs and HRs of OS(A) and PFS(B) of switch maintenance

### 5. DISCUSSION

To address the limitations of traditional AD meta-analyses of survival data, this study used simulated data to develop statistical multilevel models for the synthesis of MSTs in an

AD meta-analysis of survival data. The results show convincing outcomes of good agreement between the results from both models and real world examples.

**Table 3.** Relative changes of the multilevel model value from the gold standard value for each dataset

Dataset	Relative changes	Model 1 (%)	Model 2 (%)
M5S200	< 5%	94.4	94.4
	5%~10%	5.5	5.5
	10%~15%	0.1	0.1
	≥ 15%	0.0	0.0
M5S300	< 5%	96.4	96.4
	5%~10%	3.6	3.6
	10%~15%	0.0	0.0
	≥ 15%	0.0	0.0
M5S400	< 5%	98.8	98.9
	5%~10%	1.2	1.1
	10%~15%	0.0	0.0
	≥ 15%	0.0	0.0
M10S200	< 5%	98.6	98.4
	5%~10%	1.4	1.6
	10%~15%	0.0	0.0
	≥ 15%	0.0	0.0
M10S300	< 5%	91.9	92.3
	5%~10%	8.1	7.7
	10%~15%	0.0	0.0
	≥ 15%	0.0	0.0
M10S500	< 5%	98.8	98.7
	5%~10%	1.2	1.3
	10%~15%	0.0	0.0
	≥ 15%	0.0	0.0
M20S500	< 5%	100.0	100.0
	5%~10%	0.0	0.0
	10~15%	0.0	0.0
	≥ 15%	0.0	0.0
M20S750	< 5%	100.0	100.0
	5%~10%	0.0	0.0
	10%~15%	0.0	0.0
	≥ 15%	0.0	0.0

Note. Multilevel model 1 weighted by  $\sqrt{n_{11}}$  and  $\sqrt{n_{12}}$  for experimental and control groups, respectively. Multilevel model 1 weighted by  $SE_{11} = \sqrt{\frac{n_2}{n_1 + n_2} \times \frac{\log(HRU) - \log(HRL)}{3.92}}$  and  $SE_{12} = \sqrt{\frac{n_1}{n_1 + n_2} \times \frac{\log(HRU) - \log(HRL)}{3.92}}$  for experimental and control groups, respectively.

For time-to-event data the common distributions of survival time are exponential distribution, weibull distribution, log-normal distribution, and so on. In this study we simulated the survival time of the exponential distribution, which is considered one of the most common and basic distributions of human diseases.<sup>[27]</sup> We transferred the MSTs into a logarithm to construct the model of our meta-analysis.<sup>[15,28]</sup> The variation among studies is assumed to be the sampling error

in a fixed effect model and the deviation of true effect in a random effect model.<sup>[15,28]</sup> For the weights of the multilevel models in our synthesis of MSTs we used the square root  $n$  in each group and the standard error of  $\log HR$ , which is reasonable to represent the precision of effect size.<sup>[29]</sup>

Multilevel models provide a general framework for meta-analysis. This study incorporated MSTs from both multilevel model and meta-analysis perspectives. Since the analysis is based on a maximum likelihood estimation model, it can be expected to yield more efficient estimates than traditional meta-analyses.

A previous study evaluated retrospectively the performance of meta-analyses using MST ratios, survival rates, and HRs by comparing them with meta-analyses using IPD. The results showed that the combined HRs are mostly consistent with the IPD results. The results also showed that MST ratios and survival rates are not reasonable surrogate measures for meta-analyses of survival outcomes.<sup>[8]</sup> However, the study mainly focused on comparison of the experimental and control groups. Our study mainly focused on the synthesis of MSTs in each group, which differs from the previous research. The effect sizes are  $\log MSTs$  in each group in our study; the effect size is  $\log MSR$  between groups in the previous study. To our knowledge no other research has developed a method for synthesis of the MSTs for each group in an AD meta-analysis of survival data.

In the actual meta-analysis we pooled the HRs and their variances first. Then we calculated the combined median times to provide additional information for the pooled HRs. The effect of switch maintenance with small-molecule EGFR TKIs was smaller on OS than on PFS if we based the meta-analysis only on combined HRs (HR 0.87 for OS vs. 0.75 for PFS). However, the combined median times revealed a larger difference between the two groups in MST (12.86 vs. 11.44 months, difference of 1.42 months) than in median progression time (3.53 vs. 2.67 months, difference of 0.86 months). Our results indicate that the pooled HRs and their variances were not sufficient when conducting a meta-analysis of survival data. The median times need to be combined to provide additional information.

The strength of our study is that we first constructed statistical models for the synthesis of MSTs in an AD meta-analysis of survival data with multilevel models. We also tested the fitting effect using reasonable methods.<sup>[30,31]</sup> A second strength is that we applied these models in a meta-analysis of survival data to test the feasibility of the models.

The limitation of the study is that we considered no covariates in these models, and the results might be varied because

of the moderators. Readers should interpret the results calculated from the model carefully.

In conclusion, the multilevel statistical models for the synthesis of MSTs were developed for the past meta-analysis of survival data, which we verified to have enough fitting effect. Combined MSTs along with pooled HRs in the meta-analysis

of survival data can provide more direct evidence.

## ACKNOWLEDGEMENTS

This study was supported by the research grants of Shanghai Municipal Bureau of Health (No.20134053), Shanghai Yangfan Talents Plan Grant (No. 15YF1411100) and the Health Fields Specific Research Grant (201202012).

## REFERENCES

- [1] Pignon JP, Hill C. Meta-analyses of randomised clinical trials in oncology. *Lancet Oncol.* 2001; 2: 475-82. [http://dx.doi.org/10.1016/S1470-2045\(01\)00453-3](http://dx.doi.org/10.1016/S1470-2045(01)00453-3)
- [2] Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet.* 1993; 341: 418-22. [http://dx.doi.org/10.1016/0140-6736\(93\)93004-K](http://dx.doi.org/10.1016/0140-6736(93)93004-K)
- [3] Simmonds MC, Higgins JP, Stewart LA, *et al.* Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials.* 2005; 2: 209-17. PMID:16279144. <http://dx.doi.org/10.1191/1740774505cn0870a>
- [4] Riley RD, Simmonds MC, Look MP. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *J Clin Epidemiol.* 2007; 60: 431-9. PMID:17419953.
- [5] Simmonds MC. Statistical methodology of individual patient data. Ph.D. Thesis, University of Cambridge; 2005.
- [6] Michiels S, Piedbois P, Burdett S, *et al.* Meta-analysis when only the median survival times are known: a comparison with individual patient data results. *Int J Technol Assess Health Care.* 2005; 21: 119-25.
- [7] Whitehead A. Meta-analysis of Controlled Clinical Trials. Wiley: West Sussex; 2002.
- [8] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986; 7: 177-88. [http://dx.doi.org/10.1016/0197-2456\(86\)90046-2](http://dx.doi.org/10.1016/0197-2456(86)90046-2)
- [9] Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med.* 1998; 17: 2815-34. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19981230\)17:24<2815::AID-SIM110>3.0.CO;2-8](http://dx.doi.org/10.1002/(SICI)1097-0258(19981230)17:24<2815::AID-SIM110>3.0.CO;2-8)
- [10] Arends LR, Hunink MG, Stijnen T. Meta-analysis of summary survival curve data. *Stat Med.* 2008; 27: 4381-96. PMID:18465839. <http://dx.doi.org/10.1002/sim.3311>
- [11] Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med.* 1999; 18: 2693-708. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19991030\)18:20<2693::AID-SIM235>3.0.CO;2-V](http://dx.doi.org/10.1002/(SICI)1097-0258(19991030)18:20<2693::AID-SIM235>3.0.CO;2-V)
- [12] David Collett. Modelling Survival Data in Medical Research, Second Edition. Boca Raton: Chapman & Hall/CRC. 2003. ISBN 978-1-58488-325-8.
- [13] Zang J, Xiang C, He J. Synthesis of median survival time in meta-analysis. *Epidemiology.* 2013; 24: 337-8. PMID:23377098. <http://dx.doi.org/10.1097/EDE.0b013e318282a66c>
- [14] Zang J, Xu J, Xiang C, *et al.* Multi-level Model Synthesis of Median Survival Time in Meta-analysis. *Epidemiology.* 2015; 26: e2-3.
- [15] Borenstein M, Hedges L, Higgins J, *et al.* Computing effect sizes for Meta-analysis. 2009. Chichester: John Wiley and Sons, Ltd.
- [16] Thompson SG, Turner RM, Warn DE. Multilevel models for meta-analysis, and their application to absolute risk differences. *Stat Methods Med Res.* 2001; 10: 375-92. <http://dx.doi.org/10.1191/096228001682157616>
- [17] Miller VA, O'Connor P, Soh C, *et al.* A double-blind, randomized, phase III study of maintenance erlotinib versus placebo following nonprogression with first-line platinum-based chemotherapy in patients with advanced NSCLC. *J Clin Oncol.* 2009; 27: 18s (suppl; abstr LBA8002).
- [18] Gaafar RM, Surmont V, Scagliotti G, *et al.* A double-blind, randomized, placebo-controlled phase III intergroup study of gefitinib (G) in patients (pts) with advanced NSCLC, non-progressing after first-line platinum-based chemotherapy (EORTC 08021-ILCP 01/03). *J Clin Oncol.* 2010; 28: 15s (suppl; abstr 7518).
- [19] Cappuzzo F, Ciuleanu T, Stelmakh L, *et al.* Erlotinib as maintenance treatment in advanced non-small-cell lung cancer: a multicentre, randomized, placebo-controlled phase 3 study. *Lancet Oncol.* 2010; 11: 521-9. [http://dx.doi.org/10.1016/S1470-2045\(10\)70112-1](http://dx.doi.org/10.1016/S1470-2045(10)70112-1)
- [20] Giaccone G, Herbst RS, Manegold C, *et al.* Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: a phase III trial-INTACT 1. *J Clin Oncol.* 2004; 22: 777-84. PMID:14990632. <http://dx.doi.org/10.1200/JCO.2004.08.001>
- [21] Herbst RS, Giaccone G, Schiller JH, *et al.* Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: a phase III trial-INTACT 2. *J Clin Oncol.* 2004; 22: 785-94. PMID:14990633. <http://dx.doi.org/10.1200/JCO.2004.07.215>
- [22] Perol M, Chouaid C, Milleron BJ, *et al.* Maintenance with either gemcitabine or erlotinib versus observation with predefined second-line treatment after cisplatin-gemcitabine induction chemotherapy in advanced NSCLC: IFCT-GFPC 0502 phase III study. *J Clin Oncol.* 2010; 28: 15s (suppl; abstr 7507).
- [23] Kabbavar FF, Miller VA, Johnson BE, *et al.* Overall survival (OS) in ATLAS, a phase IIIb trial comparing bevacizumab (B) therapy with or without erlotinib (E) after completion of chemotherapy (chemo) with B for first-line treatment of locally advanced, recurrent, or metastatic non-small cell lung cancer (NSCLC). *J Clin Oncol.* 2010; 28: 15s (suppl; abstr 7526).
- [24] Gatzemeier U, Pluzanska A, Szczesna A, *et al.* Phase III study of erlotinib in combination with cisplatin and gemcitabine in advanced non-small-cell lung cancer: the Tarceva Lung Cancer Investigation Trial. *J Clin Oncol.* 2007; 25: 1545-52. PMID:17442998. <http://dx.doi.org/10.1200/JCO.2005.05.1474>
- [25] Herbst RS, Prager D, Hermann R, *et al.* TRIBUTE: a phase III trial of erlotinib hydrochloride (OSI-774) combined with carboplatin and paclitaxel chemotherapy in advanced non-small-cell lung cancer. *J Clin Oncol.* 2005; 23: 5892-9. PMID:16043829. <http://dx.doi.org/10.1200/JCO.2005.02.840>

- [26] Mok TS, Wu YL, Yu CJ, *et al.* Randomized, placebo-controlled, phase II study of sequential erlotinib and chemotherapy as first-line treatment for advanced non-small-cell lung cancer. *J Clin Oncol.* 2009; 27: 5080-7. PMID:19738125. <http://dx.doi.org/10.1200/JCO.2008.21.5541>
- [27] Ritesh S, Keshab M. *Survival analysis in clinical trials: Basics and must know areas.* Statistics. 2011; 2: 145-48.
- [28] Michael B, Hedges LV, Higgins JPT, *et al.* *Introduction to Meta-Analysis.* 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-05724-7.
- [29] Tierney JF, Stewart LA, Ghersi D, *et al.* Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials.* 2007; 8: 16. PMID:17555582. <http://dx.doi.org/10.1186/1745-6215-8-16>
- [30] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986; 1: 307-10. PMID:2868172.
- [31] Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999; 8: 135-60. <http://dx.doi.org/10.1191/096228099673819272>