THE LANCET

## Supplementary webappendix

This webappendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Parker C, Waters R, Leighton C, et al. Effect of mitoxantrone on outcome of children with first relapse of acute lymphoblastic leukaemia (ALL R3): an open-label randomised trial. *Lancet* 2010; published online Dec 4. DOI:10.1016/S0140-6736(10)62002-8.

# Parker et al Mitoxantrone improves outcome of children with first relapse of acute lymphoblastic leukaemia – results of the randomised ALL R3 trial
## Supplementary Data

**Protocol**

The protocol referred to in this paper can be obtained at the following website: http://www.medicine.manchester.ac.uk/staff/vaskarsaha The username is: R3Trial and password: Mitoxantrone. Please note that the version of the protocol posted on this website is to provide further clarifications if required with regards to this paper. Clinicians wishing to obtain a current trial protocol are requested to contact the trial coordinator, whose correspondence details are given in the paper.

**Registration and Randomisation**

We designed a web based database that permitted remote entry after an identified secure log in process using propriety software (InferMed, London UK). Individual physicians were able to track the progress of their patients. Only the trial manager, statisticians and software programmer were able to view data on all patients. Once a patient was successfully registered[1], the program verified the risk stratification. The trial statistician prepared 1:1 allocation randomisation codes stratified by risk group and country with varying blocksizes using STATA software release 7 (StataCorp, College Station, Texas, USA). Randomisation was initially by phone to the trial statistician with randomisation being allocated once patient details were complete. After initial patients, randomisation was by means of complete online registration followed by automatic allocation by computer. These practices ensured allocation concealment. To cover communication errors, an emergency list in the form of sealed envelopes was held for each country by the Trial Manager. During the trial access to the full randomisation lists were restricted to the trial statisticians, trial manager and software programmer. The trial was not blinded, so once a patient was randomised their allocation was accessible to all trial and treating staff, parents and the patient.

**Trial Chemotherapy**

*Idarubicin/Mitoxantrone*

The idarubicin dose was based on the results of the CCG 1884 trial[2] for relapsed ALL, where a dose of 12.5mg/m$^2$ was associated with unacceptable toxicity. We conducted an initial pilot study using Idarubicin 10mg/m$^2$ on 3 consecutive days. It was felt that this was too toxic and the dose reduced to 10mg/m2 on 2 consecutive days. The dose of Mitoxantrone was based on the dose used in the AML12 trial; where there has been considerable experience in the use of Mitoxantrone at 10mg/m$^2$ [2 3] (page 17, ALL R3 protocol v 4.0).

*Erwinase*

ALLR3 uses PEG-Asparaginase in Phase I and II but Erwinase in phase III. The rationale for this was two-fold. The frontline trial in UK uses PEG-Asparaginase. It was felt that subsequent use of E Coli Asparaginase, in the absence of clinical hypersensitivity symptoms, could nevertheless lead to the development of silent neutralising antibodies in relapsed patients. Thus the use of Erwinase was a blinded attempt to ensure that all patients received some form of active asparaginase during the first 3 phases of therapy. The second reason was to avoid the prolonged asparagine depletion produced by the pegylated compound in the pre-transplant block.

**Analysis**

*Cytogenetics*

Diagnostic pre-treatment bone marrow and/or blood samples were cultured and analysed in regional cytogenetic/genetic laboratories using standard methodologies. Karyotypes were collated and checked centrally by the Leukaemia Research Fund (LRF) UK Cancer Cytogenetics group (UKCCG) Karyotype Database. All karyotypes were described according to the international system for human cytogenetic nomenclature (1995). Patients were screened for *ETV6-RUNX1* fusion, *BCR-ABL* fusion and *MLL* gene rearrangements using commercially available fluorescent in situ hybridization (FISH) probes. All FISH probes were used in accordance with manufacturer's instructions. Only patients with 51-65 chromosome by conventional cytogenetic analyses, or those with a classic pattern of chromosome gain using the multiprobe-I system were classified as having high hyperdiploidy. Patients were classified as having a *MLL* translocation if an established 11q23/*MLL* translocation was seen by conventional cytogenetic analysis or a split signal pattern was seen using a breakapart *MLL* FISH

probe irrespective of the bi-banded karyotype. Poor cytogenetic subgroups with a poor outcome after relapse: *MLL* translocations, iAMP21, t(1;19)(q23;p13)/*TCF3-PBX1,* t(17;19)(q23;p13)/*TCF3-HLF*, near haploidy (<30 chromosomes), low hypodiploidy (<40 chromosomes) and t(9;22)(q34;q11.2)/*BCR-ABL1*/Philadelphia chromosome.

*Sample Size*
To achieve an 80% power to detect an increase of 10% (90% vs. 80%) in $MRD^{lo}$ at TP1 between the randomised arms, using a 2-sided 5% significance level, 219 patients were required in each arm of the study with an accrual time of 6-7 years. This sample size target was noted to be underpowered for OS & PFS, however greater accrual was not feasible given the rarity of relapsed ALL. In fact, a steady decrease in relapse rates during the life of the trial meant that the numbers recruited fell short of that anticipated.

*Closure of Randomization*
Accrual and unblinded safety data, including death and SAE listings were reported on a 6-monthly basis to an independent data monitoring committee (DMC) along with survival data not split by randomised arm. A report aggregating results over the two randomised groups was circulated to the chief and principal investigators. Trial data incorporating the DMC report was reviewed annually by a trial steering committee. Two interim analyses of outcome data were scheduled for when 150 & 300 patients had been recruited, and a stopping rule set for if the difference in total number of (all-cause) deaths reported between the two randomised arms gave a p-value of $\leq 0.001$ on a two-sided Fisher's exact test.

The first interim analysis was performed in May 2007 at which point survival differences between the randomised arms were noted. The DMC requested to view the data after a 6-month interval. At this time, while the pre-specified stopping rule boundaries had not been reached, given the differences in OS and PFS between the two arms, the DMC recommended closure of the randomisation on ethical grounds in December 2007. Supplementary table 1 shows the p-values for OS & PFS at the time of the interim analysis in November 2007 and current analysis (June 2009).

**Supplementary Table 1. Results of interim and final analyses**

| | November 2007 | June 2009 |
|---|---|---|
| 3 year OS | p=0.01 | p=0.004 |
| 3yr PFS | p=0.006 | p=0.0004 |

The mortality rates in November 2007 were 42/100 in Idarubicin and 24/102 in Mitoxantrone, p=0.007.

*Re-classification of Risk Group for Very Early isolated CNS relapse*
At the start of the trial, we originally classified patients according to a schema published by the Berlin Frankfurt Münster (BFM) group[4]. In 2005 based on our limited centre analyses of ALL R2 (the preceding trial) very early isolated extramedullary relapses were reclassified from IR to HR[5]. Patients are analysed according to the risk group they were treated as by the hospital; hence the patients in this group randomised prior to 2005 analysis remain analysed as IR. According to the therapeutic protocol, these patients were eligible for allo-SCT regardless. In addition, 2 patients inadvertently randomised in the wrong risk stratum are analysed as their true risk group since in both cases the hospital knew, and treated them as, the true risk group.

*Statistical Analysis*
Stata version 10.1 (StataCorp. 2007. Stata Statistical Software: Release 10. College Station, TX: StataCorp LP) has been used for the analysis, with the exception of analysis using competing risks methodology, for which the package R [6] was used. Analysis is by intention to treat, with the exception that the three ineligible patients randomised (Figure 2) were excluded from analysis and one patient with major protocol violations at the time of transplant was censored at transplant. For PFS and OS the primary analysis is by Kaplan-Meier plot and (unstratified) logrank test. The proportional hazards assumption was found to be acceptable for OS but a poor fit for PFS as the hazards in the two arms diverge, the treatment difference increasing over time. Alternative models to the Cox were sought but none was a better fit. To give a rough idea of magnitude of treatment effect, Cox models have been

included but should be interpreted with caution. For both OS and PFS, adjusted Cox regression was performed. MRD and SCT are not included as covariates since these are not known at baseline. *ETV6-RUNX1* status was missing in some patients, and so was imputed using multiple imputation with the ice procedure in Stata 10, under the missing at random assumption, with 10 data sets [7]. Results from the 10 data sets were combined using Rubin's rule[8]. The Cox model was repeated ignoring these covariates to demonstrate the effect of covariate adjustment (and imputation).

The Cox model for PFS was also used to evaluate separately interactions of the randomised drug with the following variables: immunophenotype, site, time from diagnosis to relapse and MRD at TP1 ($</\geq 1x10^{-4}$). For the three variables used in defining risk group, risk group is replaced in the model by all three of these variables (immunophenotype, time from diagnosis to relapse, site of relapse) since it is interactions with these terms which are of interest. Time from diagnosis to relapse is treated as a continuous variable. For MRD, only intermediate risk patients with bone marrow or combined relapse were included; the only terms fitted were MRD, randomised arm and their interaction due to reduced event numbers.

Logistic regression is used to explore the relationship between the drugs and MRD levels at day TP1. Only intermediate risk patients with bone marrow or combined relapse were included in analysis; patients with indeterminate or missing result (including all induction deaths and some refractory patients) were excluded. The model was fitted with and without the covariates described for Cox modelling above, except that risk group was not fitted.

The number of toxicities at grade $\geq 3$ per patient was modelled using Poisson regression, and the incidence rate ratio (IRR) presented with 95% confidence intervals. The primary comparison for toxicity uses all toxicity data over the entire duration of the trial; this is then repeated for each phase separately. No covariate adjustment was used. The IRR and confidence interval are presented for toxicities overall, by phase and by system. Toxicities post-SCT are not counted in whole-trial toxicities because they were not graded; post-SCT incidence shown is for all toxicities.

To formally assess whether differences seen between randomised drug tend to be due to an influence on disease progression and/or side effects of treatment, a competing risks model (Fine & Gray) was used splitting PFS events into disease-related events (progression, relapse, disease-related deaths) and treatment-related events (treatment-related deaths, second malignancies). Two deaths in remission clearly neither related to disease nor to treatment were treated as a third competing risk. Gray's test was used with a cumulative incidence plot, since the Fine & Gray model relies on proportional hazards but Gray's test does not.

**Results**

Five Idarubicin and 8 Mitoxantrone patients failed induction i.e. were not in remission and were withdrawn at the end of phase I. In addition, 3 Mitoxantrone patients withdrew during induction therapy. One patient died prior to starting treatment and another left the country. One patient was withdrawn by their physician.

**Randomised outcome is not influenced by study group**

**Supplementary Table 2**

| | All countries | | UK only | |
|---|---|---|---|---|
| | **Idarubicin** | **Mitoxantrone** | **Idarubicin** | **Mitoxantrone** |
| **Patients** | 109 | 103 | 98 | 96 |
| **Survival rate at 3 years** | 45.2% (34.5%, 55.3%) | 69.0% (58.5%, 77.3%) | 45.2% (34.3%, 55.5%) | 68.4% (57.7%, 76.9%) |
| **Survival rate at 5 years** | 43.5% (32.7%, 53.7%) | 60.9% (46.8%, 72.3%) | 43.5% (32.6%, 54.0%) | 60.4% (46.2%, 71.9%) |

Median follow up by the reverse Kaplan-Meier method:
UK 44 months (95% CI 39, 47)
Australia / New Zealand 16 months (95% CI 13, 18)

**Randomised outcome is not affected by other covariates**

**Supplementary Table 3**

| | Adjusted analysis | | Unadjusted analysis | |
|---|---|---|---|---|
| | **Hazard ratio (95% CI)** | **P-value** | **Hazard ratio (95% CI)** | **P-value** |
| **PFS** | 0.54 (0.36, 0.82) | 0.003 | 0.49 (0.33, 0.73) | 0.001 |
| **OS** | 0.56 (0.36, .087) | 0.01 | 0.54 (0.35, 0.83) | 0.005 |

To adjust for potential differential distribution of factors that could influence outcome, an adjusted Cox regression analysis was performed. Covariates were adjusted for were the factors comprising risk group, which was used to stratify randomisation (duration of CR1, immunophenotype and site of relapse) and factors expected to be prognostic (ETV6-RUNX1, age and gender). As shown in supplementary table 3, adjusting or not, for these covariates, Mitoxantrone almost halves the hazard of an event at any given time point for both PFS and OS. The results remain unchanged if analysis is restricted to UK patients only (Supplementary Table 4).

**Supplementary Table 4**

| | Adjusted analysis | | Unadjusted analysis | |
|---|---|---|---|---|
| | **Hazard ratio (95% CI)** | **P-value** | **Hazard ratio (95% CI)** | **P-value** |
| **All countries** | 0.56 (0.36, 0.87) | 0.01 | 0.54 (0.35, 0.83) | 0.005 |
| **UK only** | 0.60 (0.38, 0.94) | 0.03 | 0.55 (0.36, 0.86) | 0.009 |

**Supplementary Table 5**

| Interaction | Level | Hazard ratio (95% CI) | P-value |
|---|---|---|---|
| **MRD** | | | 0.73 |
| | $<1x0^{-4}$ | 0.51 (0.12, 2.13) | |
| | $\geq 1x10^{-4}$ | 0.36 (0.11, 1.22) | |
| **Immunophenotype** | | | 0.53 |
| | B-cell | 0.46 (0.30, 0.73) | |
| | T-cell | 0.68 (0.23, 1.96) | |
| **Relapse site** | | | 0.22 |
| | Isolated BM | 0.68 (0.38, 1.19) | |
| | Combined | 0.24 (0.08, 0.68) | |
| | Isolated EM | 0.42 (0.19, 0.94) | |
| **Time to relapse** | | | 0.54 |
| | 12 Months | 0.57 (0.30, 1.08) | |
| | 24 Months | 0.53 (0.28, 0.99) | |

A Cox model was used to examine the interactions of the various pre-specified prognostic factors with the randomised outcome. Only IR patients with bone marrow involvement are included when assessing interaction with MRD. As shown in supplementary table 5, the variation seen in the hazard ratios for idarubicin to mitoxantrone in subgroups is not statistically significant. Thus none of the variables showed a significant interaction with drug effect.

**Randomised outcome is not due to differences in compliance**

It is common for children with relapsed ALL, undergoing intensive treatment to miss blocks of therapy, due to subjective decisions made by physicians and/or family. For example they may decide that a child will not tolerate a particular phase of therapy given their past clinical history. To ensure this did not affect the results of the study a sensitivity analysis was carried out using only patients fully compliant with treatment, for PFS only. Patients recorded to have missed any scheduled therapy or where doses were altered are included in this category. Patients with unknown compliance due to missing data are assumed to be compliant, and included in this analysis; only those known not to be compliant are excluded. This leaves 91 patients each for idarubicin and mitoxantrone. Results are shown for all countries and not repeated for UK patients alone.

**Supplementary Table 6**

| | Main analysis | | Sensitivity analysis | |
|---|---|---|---|---|
| | **Idarubicin** | **Mitoxantrone** | **Idarubicin** | **Mitoxantrone** |
| **Patients** | 109 | 103 | 91 | 91 |
| **Progression free rate at 3 years** | 35.9% (25.9%, 45.9%) | 64.6% (54.2%, 73.2%) | 34.6% (23.8%, 45.6%) | 62.2% (51.1%, 71.6%) |
| **Progression free rate at 5 years** | 32.2% (22.3%, 42.6%) | 60.7% (49.4%, 70.2%) | 29.6% (18.9%, 41.2%) | 59.9% (48.2%, 69.8%) |

The results of Cox regression show a similar hazard ratio for mitoxantrone:idarubicin to that in the primary analysis of 0.56 (0.36, 0.88) with a p value of 0.011

**Randomised outcome is unrelated to MRD levels at TP1**

Data for IR MRD has been presented in Table 3. MRD at TP1 cannot be evaluated in patients with isolated extramedullary disease, which includes all SR patients. As TP1 MRD was not required for treatment decisions in HR, this is mostly unavailable. Of 8 patients in the HR group, where MRD was assessed, 7 were $MRD^{hi}$ and 1 was $MRD^{lo}$.

**Supplementary Table 7 MRD results at TP1 and TP2**

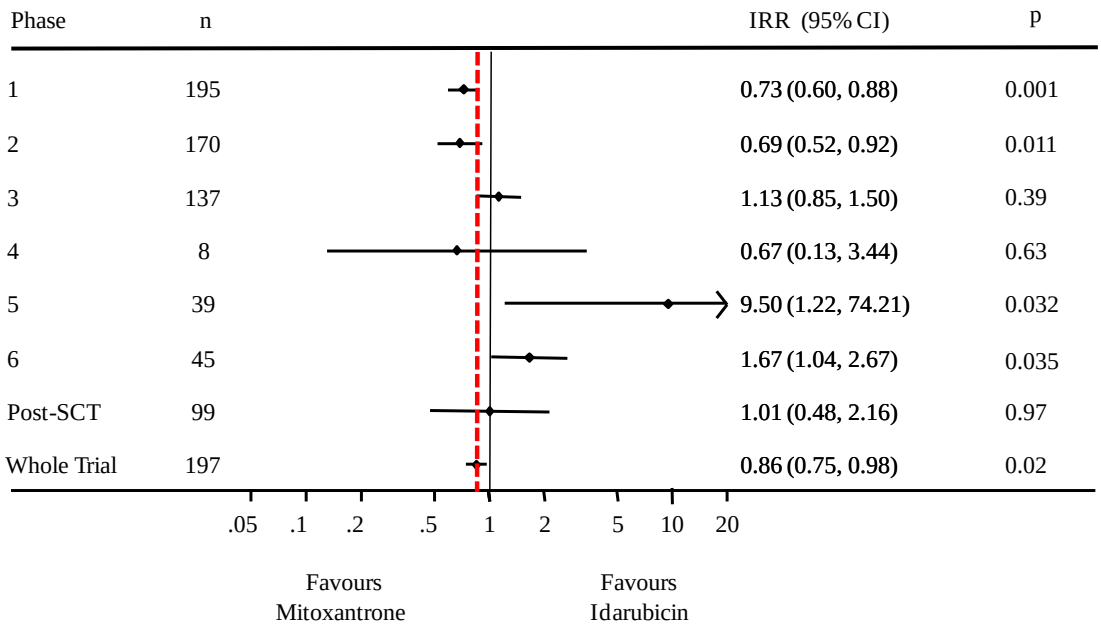
| Category | Idarubicin | | Mitoxantrone | |
|---|---|---|---|---|
| Extramedullary | 29 | | 14 | |
| TP1 | **HR** | **IR** | **HR** | **IR** |
| Indeterminate | 15 | 12 | 12 | 18 |
| $MRD^{lo}$ | 0 | 16 | 1 | 17 |
| $MRD^{hi}$ | 6 | 21 | 1 | 24 |
| TP2 | | | | |
| $<10^{-3}$ | 6 | 22 | 3 | 17 |
| $>10^{-3}$ | 5 | 4 | 2 | 4 |

**Randomised outcome is related to differences in disease clearance and not in toxicity**

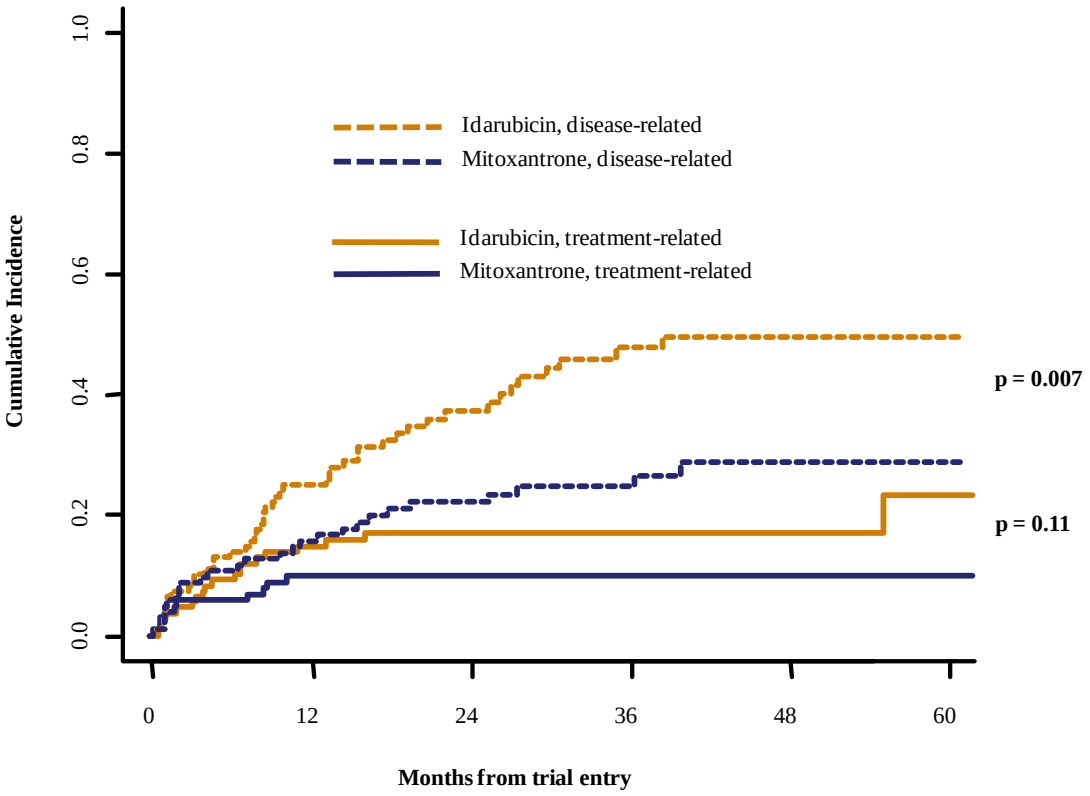
**Supplementary Table 8 Incidence of toxicities**

| Subgroup | Patients giving toxicity data | | Median number of toxicities at grade 3+ (IQR) | | P-value |
|---|---|---|---|---|---|
| | Idarubicin | Mitoxantrone | Idarubicin | Mitoxantrone | |
| Over whole trial | 102 | 95 | 4 (1, 8) | 3 (1, 7) | 0.02 |
| I (induction) | 101 | 94 | 1 (0, 3) | 0 (0, 2) | 0.001 |
| II (consolidation) | 90 | 80 | 1 (0, 2) | 1 (0, 1) | 0.011 |
| III (intensification) | 72 | 65 | 1 (0, 2) | 1 (0, 2) | 0.39 |
| IV (FLAD) | 5 | 3 | 1 (0, 1) | 0 (0, 2) | 0.63 |
| V (Interim Maintenance) | 19 | 20 | 0 (0, 0) | 0 (0, 1) | 0.032 |
| VI (Maintenance) | 22 | 23 | 0 (0, 2) | 0 (0, 2) | 0.035 |
| Post-SCT | 36 | 33 | 0 (0, 1) | 0 (0, 1) | 0.97 |
| Other toxicities | 102 | 95 | 1 (0, 3) | 0 (0, 2) | 0.032 |
| Liver toxicities | 102 | 95 | 1 (0, 2) | 0 (0, 1) | 0.14 |
| Infections | 102 | 95 | 0 (0, 2) | 1 (0, 2) | 0.91 |
| GI toxicities | 102 | 95 | 0 (0, 1) | 0 (0,1) | 0.56 |

**Supplementary Figure 1 Forest plot showing toxicity of the randomised drugs during the different phases of treatment**




**Supplementary Figure 2 Cumulative incidence of disease- and treatment-related events by induction.** The graph is curtailed at 5 years due to low numbers of patients observed beyond that point.



**References**


1. Tartarone A, Sirotova Z, Aieta M, Lelli G. Salvage treatment with epirubicin and/or paclitaxel in metastatic breast cancer patients relapsed after high-dose chemotherapy with peripheral blood progenitor cells. Tumori 2001;87:134-7.
2. Leahey A, Bunin N, Belasco J, et al. Novel Multiagent chemotherapy for bone marrow relapse of pediatric acute lymphoblastic leukemia. Med Pediatr Oncol 2000;34:313-8.
3. Gibson BE, Wheatley K, Hann IM, et al. Treatment strategy and long-term results in paediatric patients treated in consecutive UK AML trials. Leukemia 2005;19:2130-8.
4. Borgmann A, von Stackelberg A, Hartmann R, et al. Unrelated donor stem cell transplantation compared with chemotherapy for children with acute lymphoblastic leukemia in a second remission: a matched-pair analysis. Blood 2003;101:3835-9.
5. Roy A, Cargill A, Love S, et al. Outcome after first relapse in childhood acute lymphoblastic leukaemia - lessons from the United Kingdom R2 trial. British Journal of Haematology 2005;130:67-75.
6. Team RDC. R: A Language and Environment for Statistical Computing. In. Vienna: R Foundation for Statistical Computing; 2010.
7. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in medicine 1999;18:681-94.
8. Little RJA, Rubin D. Statistical Analysis with Missing Data. New York: John Wiley and Sons; 1987.