Supplementary Material

Supplement to CMV viral load kinetics as surrogate endpoints after allogeneic transplantation

Table of Contents

Supplementary Methods

Goodrich et al. NEJM 1991 Early Treatment Clinical Trial Methods

Complete study methods are available in the original Goodrich *et al.* publication,[1] but additional information not available in the main text of the manuscript is provided here. Inclusion criteria were age two years of age or older, undergoing allogeneic bone marrow transplant for hematologic malignancy, CMV seropositivity before transplantation or receiving marrow from a CMV seropositive donor. Viral surveillance cultures from blood, urine, and throat swabs were collected weekly. In addition, broncheoalveolar lavage (BAL) cultures were collected from a subset of these patients participating in a CMV pneumonia prevention study on day 35 post-transplant. Specimens were cultured using both centrifugation culture and conventional viral culture. If any of the surveillance or BAL cultures was positive prior to day 80 following transplant, patients whose transplants had engrafted (absolute neutrophil count of 500 x 10^6 per milliliter for at least two days) were randomized to receive ganciclovir or placebo.
Patients were excluded if serum creatinine was greater than 220 mmol per liter, if they had symptomatic CMV disease, if they had received therapy for CMV or any other investigational antiviral within seven days of enrollment, or if they had sensitivity to acyclovir or ganciclovir. Randomization groups were stratified based on whether patients had developed acute graft-versus-host disease prior to randomization.

Patients in the treatment group received 5 mg per kg of intravenous ganciclovir twice daily for seven days followed by once daily dosing for the remainder of the 100-day period following transplant or until the patient left Seattle if earlier. Study staff remained in contact with patients and their providers to collect information regarding CMV infection, disease, and death for 180 days after transplantation. The Food and Drug Administration and the FHCRC institutional review board approved the protocol. All patients or their legal guardians provided informed consent.

Goodrich et al. AIM 1993 Prophylaxis Clinical Trial Methods

The double-blind, placebo-controlled, randomized clinical trial for ganciclovir prophylaxis to prevent CMV infection after allogeneic bone marrow transplant was conducted at the Fred Hutch between November 1990 and August 1991. Complete study methods are available in the original Goodrich *et al.* publication[2]. To summarize, prior to transplantation, study staff enrolled all CMV seropositive patients two years of age or older undergoing allogeneic bone marrow transplant for hematologic malignancy requiring total body irradiation or busulfan-cyclophosphamide (myeloablative conditioning regimens). Patients were randomized at marrow engraftment (absolute neutrophil count of 750 per microliter or greater for two consecutive days) to receive ganciclovir or placebo at a dose of 5 mg/kg twice daily for 5 days and then daily through day 100 after HCT or until a study endpoint was reached. Patients were excluded if serum creatinine was greater than 220 mmol per liter, if they received a T cell-depleted transplant, were allergic to acyclovir or ganciclovir, had received a marrow transplant in the last 6 months, or had documented CMV excretion before randomization.

Viral surveillance cultures from blood, urine, and throat swabs were collected weekly. Specimens were cultured using both centrifugation culture and conventional viral culture. Patients were considered to have reached a primary endpoint if they developed CMV infection, defined as a positive viral culture from the throat, urine, or blood, or if they developed neutropenia, defined as an absolute neutrophil count of 750 cells per microliter or fewer for two consecutive days. CMV disease confirmed by biopsy or culture and mortality were considered secondary endpoints. All

patients who developed CMV infection or disease prior to 100 days post-transplant were removed from the study and treated with ganciclovir. Study staff remained in contact with patients and their providers to collect information regarding CMV disease and death for 180 days after transplantation. The Food and Drug Administration and the FHCRC institutional review board approved the protocol. All patients or their legal guardians provided informed consent.

Extended Clinical Outcome Analysis

Study materials from the original Goodrich *et al.* clinical trials were unavailable, as the sponsoring pharmaceutical company no longer exists. Thus, we reconstructed patient demographics, laboratory, and clinical data, including time-to-event data for both CMV disease and overall mortality, from review of the Fred Hutchinson Cancer Research Center (Fred Hutch) research database and chart review of scanned medical records. We extended the time-to-event data for both CMV disease and mortality to twenty years by reviewing records from the Long Term Follow Up (LTFU) Clinic at Fred Hutch. The LTFU maintains contact with patients and their providers and sends annual surveys that specifically address CMV infection, CMV disease, and death for patients who return to their local providers for ongoing care following transplant.

CMV DNA Viral Load Testing

The University of Washington Molecular Virology Laboratory performed CMV DNA PCR testing using a laboratory-developed assay[3]. The assay's limit of quantification is 71.4 IU/mL; the limit of detection is 35.7 IU/mL. Because we are unable to determine precisely viral loads between 35.7 IU/mL and 71.4 IU/ml, we used the median between the limit of detection and limit of quantification, 53.6 IU/mL, for any viral loads that were determined by raw testing to be between those values. For values below the limit of detection, we used the median between 0 IU/mL and 35.7 IU/mL, 17.9 IU/mL.

Statistical Analysis

Survival and Cumulative Incidence Analysis

Survival and cumulative incidence of CMV disease and first event of CMV disease or death were determined using Kaplan-Meier and Aalen-Johnson methods, respectively[4,5] in R (version 3.5.0)[6]. Survival distributions and times to the composite endpoint CMV disease or death were compared using the log-rank test. Cumulative incidence distributions for CMV disease with death as a competing risk were compared using Gray's test[5]. Throughout the analysis, differences were considered significant when p-values were less than 0.05 unless otherwise indicated. All p-values were two-sided, and no adjustments were made for multiple hypothesis testing.

Viral Load Kinetics

In order to define viral load kinetics at discrete time points, we determined a baseline viral load and binned subsequent viral loads into weekly intervals after randomization. Baseline viral loads were chosen on the day of randomization if available. Otherwise, the nearest day either before or within three days after randomization was chosen. If no samples were available within three days of randomization, the closest sample collected prior to randomization was used. For patients who had samples collected within three days of randomization but were equally close to the day of randomization (i.e. two days before and two days after), the sample collected prior to randomization was chosen. To determine weekly viral load values after randomization, samples collected within three days (before or after) of each week (i.e. day 7, day 14, day 21, etc.) were assigned to that week. If there were two samples collected in that time frame, the sample collected

closest to the week was included. If two data samples were collected equally far from the week, an average viral load was taken and then log converted. Viral load data collected after the first event of CMV disease or death were removed from the analysis.

Viral load was defined as the log 10-converted viral load measured in IU/mL. Change in viral load was calculated by taking the maximum of weekly change in viral load, calculated by subtracting week 1 through week 5 viral loads from the baseline viral load (log 10 IU/mL), e.g. $\Delta VL_1 = VL_{rand} - VL_1$. Peak viral load was defined as the highest log-10 converted viral load (IU/mL) measured from week 1 to week 5. Mean VL was calculated as the average of viral load from week 1 to week 5. Percentage of positive viral loads (% Pos) was defined as the number of weekly viral loads with measurements at or above the limit of detection divided by the number of weekly viral loads available, multiplied by 100.

Validation of Surrogate Markers

Super Learner Machine Learning Analysis

Super Learner models were fit in the R package 'SuperLearner,' version 2.0-26[7]. Super Learner uses a library of candidate prediction algorithms. Cross-validation is used to determine the weighted combination of these algorithms that maximizes a cross-validated criterion. Due to the relatively small sample size, we used leave-one-out cross-validation: each algorithm was fit using all the data except the data for one patient. Then, that algorithm was used to predict the probability of the clinical outcome in the "left-out" patient. The procedure was repeated until the algorithm was fit and predictions made for all patients with each algorithm. The predictions from each algorithm in the library were then weighted to maximize the cross-validated area under the receiver operating characteristic curve (cvAUROC) using the R packages 'cvAUC,' version 1.1.0, and 'ROCR,'version 1.0-7[8,9]. The weighted combination of algorithms is called the Super Learner.

We used a library of six relatively simple learning algorithms: 'SL.glm,' a logistic regression model, 'SL.glm.interaction', a logistic regression model allowing for interaction terms, 'SL.step', a logistic regression model with step-wise model selection by AIC (Akaike information criteria), 'SL.bayesglm,' a Bayesian logistic regression model, 'SL.glmnet', a penalized logistic regression model, 'SL.xgboost,' a gradient-boosted tree algorithm, and 'SL.mean,' the simple mean model with no covariates. The 'arm[10],' version 1.10-1, 'glmnet[11],' version 2.0-16, 'lme4[12],' version 1.1-13, 'nnls[13],' version 1.4, and 'xgboost[14],' version 0.90.0.2 packages were used by 'SuperLearner' to fit the models.

We fit Super Learner prediction models with baseline covariates and viral load kinetics where defined for each week after randomization (weeks 1 through 5) on the clinical endpoints of CMV disease and first event of CMV disease or death by weeks 8, 24, and 48 after randomization. The baseline covariate information included acute graft versus host disease, donor CMV serostatus, and baseline viral load. We fit Super Learner models for the placebo group alone, the ganciclovir group alone, and the combined treatment groups and determined the cv-AUCs for each algorithm[8,9]. For the prophylaxis trial, we fit Super Learner models with baseline covariates and viral load kinetics on the clinical endpoint of CMV disease by week 24 after randomization.

In addition to the statistical packages referenced above, the 'lubridate'[15] package was used to convert dates to R format, and 'dplyr'[16] package was used to subset and organize data in the analysis. Plots were created using the 'ggplot2' and 'cowplot' packages[17,18].

Percentage of Treatment Effect Captured by the Candidate Surrogate

Kobayashi and Kuroki identified important limitations of previous methods for quantifying the proportion of treatment effect explained by potential surrogate markers: (1) estimates often fall outside the range of 0 to 100%, resulting in lack of interpretability; (2) methods lack defined thresholds with which to judge candidate surrogates based on the measure[19].

To overcome limitations of previous methods, Kobayashi and Kuroki proposed a new measure for quantitatively evaluating candidate surrogates called "the proportion of treatment effect captured by candidate surrogate endpoints" (PCS)[19]. Their method divides treatment effect into the portion captured by the candidate surrogate (CP) and the portion not captured by the candidate surrogate (NCP). They define the PCS as

$$PCS = \frac{CP^2}{CP^2 + NCP^2}, \#(1)$$

This measure is guaranteed to fall between 0 and 100% provided the treatment effect falls between 0 and 100%.

The CP and NCP may take different forms depending on the measure being used for treatment effect (TE), but in all cases, the CP and NCP must satisfy the following:

(a) the sum of the CP and NCP is equal to the TE
(b) the CP and NCP are quantities of the same type and unit as TE, and
(c) $NCP = 0$ holds when the candidate surrogate is a perfect surrogate, and $CP = 0$ holds when the candidate surrogate is a useless surrogate endpoint[19].

We define X , S , and Y such that X indicates treatment group ($X = x_0$ if the patient received placebo, or $X = x_1$ if the patient received ganciclovir); S represents the candidate surrogate (the VL kinetics); and Y represents the clinical outcome ($Y = 1$ if a patient has disease or the composite outcome of disease or death prior to a specified time, and $Y = 0$ if the patient does not have the specified clinical outcome). We define D_x as the domain of X, with a similar notation for other domains. The observed values of the variables X, S, and Y are denoted x, s, and y, respectively ($x \in D_x, s \in D_s, y \in D_y$). Also, $\Pr(S = s|X = x) = \Pr(S|x)$ represents conditional probability of $S = s$ given $X = x$. $E(Y|x,s)$ indicates a conditional expectation of Y given $(X,S) = (x,s)$. This notation is the same as the notation used by Kobayashi and Kuroki[19].

In our study, PCS was calculated by defining the treatment effect as the relative risk difference, i.e.

$$TE = E(Y|X = x_0) - E(Y|X = x_1). \#(2)$$

We chose to order this equation as the expected value in the placebo group minus the treatment group because we expect more clinical outcomes to occur in the placebo group, likely resulting in a positive value for TE.

Then, we define CP and NCP as

$$CP = \sum_{D_s} E(Y|x_0, s)\{\Pr(s|x_0) - \Pr(s|x_1)\}; \#(3)$$

$$NCP = \sum_{D_s} \{E(Y|x_0, s) - E(Y|x_1, s)\} \Pr(s|x_1) . \#(4)$$

To estimate these quantities, first we fit a logistic regression model on the full data set, including all patients from both the placebo and ganciclovir treatment groups, using the candidate surrogate (viral load kinetic marker), X_{VL}, and treatment group assignment, X_{GCV}, as predictors of the log odds of a clinical outcome, $\log\left(\frac{p}{1-p}\right)$, where $p = \Pr(Y = 1|X = x)$. X_{VL} assumes the units of the VL kinetics as described above. $X_{GCV} = 0$ for patients in the placebo group, and $X_{GCV} = 1$ for patients in the ganciclovir group. We use the following notation to refer to the logistic regression model:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_{GCV}X_{GCV} + \beta_{VL}X_{VL}\,.\#(5)$$

Solving this equation for the probability of a clinical event, p, yields

$$p = \frac{e^{(\beta_0+\beta_{GCV}X_{GCV}+\beta_{VL}X_{VL})}}{1+e^{(\beta_0+\beta_{GCV}X_{GCV}+\beta_{VL}X_{VL})}}.\#(6)$$

CP was estimated by calculating p for each patient using only the value of the surrogate (i.e. $X_{G\square V} = 0$) such that

$$p = \frac{e^{(\beta_0+\beta_{VL}X_{VL})}}{1+e^{(\beta_0+\beta_{VL}X_{VL})}},\#(7)$$

and assigning $E(Y|x_0, s)$ the value of p:

$$E(Y|x_0, s) = \frac{e^{(\beta_0+\beta_{VL}X_{VL})}}{1+e^{(\beta_0+\beta_{VL}X_{VL})}}.\#(8)$$

Thus, CP (the portion of the treatment effect captured by the candidate surrogate) is essentially the estimated percentage of patients predicted to develop the clinical outcome in the placebo group minus the percentage of patients predicted to develop the clinical outcome in the treatment group using only the surrogate value to make the prediction.

The NCP was estimated using the same model but including the X_{GCV} term to calculate the expected values, where $X_{GCV} = 0$ for patients in the placebo group and $X_{GCV} = 1$ for patients in the ganciclovir group:

$$E(Y|x_0, s) = \frac{e^{(\beta_0+\beta_{VL}X_{VL})}}{1+e^{(\beta_0+\beta_{VL}X_{VL})}},\#(9)$$

and

$$E(Y|x_1, s) = \frac{e^{(\beta_0+\beta_{GCV}+\beta_{VL}X_{VL})}}{1+e^{(\beta_0+\beta_{GCV}+\beta_{VL}X_{VL})}}.\#(10)$$

Thus, NCP (the portion of treatment effect not captured by the candidate surrogate), is the number of patients predicted to develop the clinical outcome based on the surrogate value and treatment assignment in the placebo group minus the number in the treatment group. The logistic regression model was fit and calculations performed using the glm function in base R[6].

Supplementary Results

RCT Cohort Demographics, Sample Availability, and Clinical Outcomes

Study Populations

In the early treatment trial, a total of 347 patients two years of age and older who were either themselves CMV seropositive or who had CMV seropositive donors underwent allogeneic bone marrow transplantation at Fred Hutch during the study period. Of these, 66 were excluded from further screening due to lack of engraftment, elevated serum creatinine, receipt of an anti-CMV drug or investigational antiviral in the previous seven days. The remaining 281 patients met

general eligibility criteria and were screened for virus excretion. Eighty-seven patients (31%) had no positive CMV cultures; 67 patients (24%) died, relapsed, or were discharged without a positive CMV culture; 18 patients (6%) developed CMV disease at the same time their surveillance cultures were positive; 17 patients (6%) developed CMV disease without any prior positive surveillance cultures; 20 patients declined participation; 72 (26%) were enrolled in the study with 35 randomized to receive placebo and 37 randomized to receive ganciclovir. The two groups were well balanced (Supplementary Table 1) except that more patients in the ganciclovir group received marrow from HLA-mismatched donors (11 versus 3). The mean time to randomization in the ganciclovir group was 54 days and in the placebo group was 48 days after HCT.

In the prophylaxis trial, 114 CMV seropositive patients met general eligibility criteria. 93 were enrolled into the study before transplant, whereas three refused entry, three received T cell-depleted grafts, and 15 were enrolled in other protocols that precluded participation in this study. Between enrollment and engraftment, 23 patients became ineligible due to renal failure, hematologic relapse, refusal, engraftment failure, and positive CMV cultures. Seventy patients were randomized and received study drug. However, when interim analysis showed a large difference in primary endpoints between the treatment and placebo arms, the study was stopped. At that time, five of the 70 patients had not reached an endpoint. All five of these patients had been randomized fewer than 2 weeks prior to the result of the interim analysis and were not included in the final study analysis. One additional patient had a positive CMV culture from a sample that had been collected prior to randomization and had received only two doses of study drug when the culture resulted. This patient was withdrawn from the study, leaving a total of 64 patients included in the final analysis.

The two groups were well balanced (Supplementary Table 2). Patients ranged in age from 3 to 56 years old. All patients received allogeneic bone marrow transplantation for hematologic malignancies, including acute myelogenous leukemia, acute lymphoblastic leukemia, chronic myelogenous leukemia, myelodysplastic syndrome, lymphoma, and multiple myeloma. All conditioning was myeloablative with some combination of total body irradiation, busulfan, and cyclophosphamide. The mean time to randomization was 25 days after HCT in the ganciclovir group and 24 days in the placebo group.

Sample Availability and Post-Randomization Clinical Outcomes

In the early treatment trial, all 72 patients had baseline viral load data available. 63 patients had a baseline randomization sample within three days of randomization; for five patients, baseline was four days prior to randomization; for the remaining four patients, baseline was 6 days, 7 days, 9 days, and 10 days prior to randomization. Because not all patients had samples available for each weekly interval after randomization and because several patients met the primary endpoint of CMV disease or died soon after randomization, the viral load data set for each week after treatment randomization is smaller and contains fewer clinical endpoints than the baseline viral load set (n = 72) and the summary viral load kinetic data set (n = 65) (Supplementary Table 3).

In the prophylaxis trial, all 64 patients had baseline viral load data available. 63 had baseline viral load samples available at or prior to randomization. One patient had a baseline viral load sample collected on day 3 after randomization. 18 patients had samples collected on the day of randomization; 38 had baseline samples collected between day one and day 4 prior to randomization; the remaining 7 patients had baseline samples available at 5 days, 6 days, 8 days, and 9 days prior to randomization. To determine weekly viral load values after randomization, samples collected within 3 days (before or after) of each week (i.e. day 7, day 14, day 21, etc.) were assigned to that week. If there were two samples collected in that time frame, the sample collected closest to the week was included. If two data samples were collected equally far from the

week, an average of log10-converted viral loads was taken. Viral samples collected after the first event of CMV disease or death were removed from the analysis. Following the removal of these samples, there were 54 samples tested at week 1; 50 at week 2; 52 at week 3; 50 at week 4; 45 at week 5; 46 at week 6; 44 at week 7; 40 at week 8; 41 at week 9; 25 at week 10; 11 at week 11; 2 at week 12; and 1 at week 14.

Effect of Ganciclovir on CMV Disease and Mortality by the Original Study Endpoints

In the early treatment trial, the primary endpoint, CMV disease in the first 100 days after transplant, developed in 15 patients (43%) in the placebo group and two patients (5%) in the ganciclovir group. In the placebo group six patients were diagnosed with CMV enteritis and nine were diagnosed with CMV pneumonia. In the ganciclovir group, one patient was diagnosed with a CMV-positive lingual ulcer, and one patient was diagnosed with CMV pneumonia. Overall mortality was also lower in the ganciclovir group. Six patients (16%) in the placebo group died from CMV disease by day 100 and one died (3%) from leukemic relapse in the ganciclovir group.

Although the study ended at day 100 post-transplant, Goodrich *et al.* were able to follow patients until day 180 and found that one additional patient in the placebo group had developed CMV disease as opposed to five in the ganciclovir group. The reduction in CMV disease remained statistically different ($p = 0.006$). Overall mortality and the combined endpoint of CMV disease or death were also lower in the ganciclovir group 180 days after transplant ($p = 0.03$; $p = 0.004$).

In the prophylaxis trial, the primary endpoint of CMV infection (by viral culture) in the first 100 days after transplant, developed in 14 patients (45%) in the placebo group and one patient (3%) in the ganciclovir group. The primary endpoint of neutropenia occurred in no patients in the placebo group and in 10 patients (30%) in the ganciclovir group.

The clinical outcomes of interest in our surrogate analysis were the secondary endpoints for the original study: tissue-invasive CMV disease and overall mortality. CMV disease developed in nine patients (29%) in the placebo arm and none of the patients in the ganciclovir arm in the first 100 days after HCT. The cumulative incidence of CMV disease by 100 days after HCT was significantly lower in the ganciclovir group by Gray's test with death treated as a competing risk ($p < 0.001$, Supplementary Figure 5A). In the placebo group three patients developed CMV pneumonia; five developed CMV gastroenteritis; and one developed both CMV pneumonia and gastroenteritis. Four of these had a surveillance culture positive prior to disease being diagnosed, whereas 5 had no positive surveillance cultures prior to diagnosis. Overall mortality was not significantly different by 100 days post-HCT (log-rank test, $p = 0.25$) with 6 dying in the placebo group and 4 dying in the ganciclovir group (Supplementary Figure 5A). Aside from two in the placebo group who died from CMV pneumonia, all others died from relapse, bacterial or fungal infection, or transplant-related-mortality. The combined endpoint of CMV disease or death was lower in the ganciclovir group (Supplementary Figure 5A, $p <0.004$).

Although the study ended at day 100, Goodrich *et al.* were able to follow patients until day 180 and found that one additional patient in the placebo group developed CMV disease as opposed to 3 in the ganciclovir group. The reduction in CMV disease remained statistically different (Supplementary Figure 5A, $p = 0.03$), but overall mortality and the combined endpoint of CMV disease or death were similar in the two groups (Supplementary Figure 5A).

Effect of Ganciclovir on CMV Disease and Mortality in Extended Follow Up Analyses

Through extended chart review, we found in the early treatment trial that at 1 year, 3 years, and even 20 years after transplantation, the difference between CMV disease incidence in the

ganciclovir group remained significantly lower (1 year, p = 0.02; 3 years, p = 0.02; 20 year, p = 0.01). Mortality and the combined endpoint of CMV disease or death remained lower in the ganciclovir group at 1 yr (p = 0.01; p = 0.006), 3 years (p = 0.04; p = 0.01), and 20 years (p = 0.11; p = 0.02) though the trend in mortality alone lost significance at 20 years.

In the prophylaxis trial, we found that there was a trend toward a lower incidence of CMV disease in the ganciclovir group through 20 years after transplantation, but this difference was not significant (Supplementary Figure 5B). Survival and the cumulative incidence of the composite endpoint of CMV disease and death were similar throughout the 20 year follow up (Supplementary Figure 5B).

Effect of Ganciclovir on CMV Disease and Death Measuring Outcomes from Randomization

In the early treatment trial, CMV disease was lower in the ganciclovir group by 100 and 180 days after randomization (p = 0.04; p = 0.04). Overall mortality and the combined endpoint of CMV disease or death were also lower in the ganciclovir group by 100 and 180 days post-randomization; the difference in overall mortality was not significant by 100 days (Supplementary Figure 1). In an extended analysis, we found that CMV disease and the combined endpoint of CMV disease or death were significantly less frequent in the ganciclovir group by 1, 3, 10, and 20 years after randomization. Overall mortality was lower out to 20 years from randomization. The difference was significant by 1 and 3 years (Supplementary Figure 1).

In the prophylaxis trial, the cumulative incidences of CMV disease were lower in the ganciclovir group by 14 and 24 weeks, but the difference lost significance at 48 weeks by Gray's test. Mortality rates were similar in the treatment and placebo groups at all times points by the log-rank test. The cumulative incidence of the composite endpoint of CMV disease or death was lower in the ganciclovir group by week 14 but not by weeks 24 and 48.

Because a surrogate endpoint for treatment effect on a clinical outcome can be validated only in the case of a successful intervention, we analyzed only CMV disease by 14 and 24 weeks and the composite endpoint of CMV disease or death by 14 weeks in the validation procedure using the Prentice Criteria and the treatment effect captured by viral kinetics.

References

1 Goodrich JM, Mori M, Gleaves CA, *et al.* Early treatment with ganciclovir to prevent cytomegalovirus disease after allogeneic bone marrow transplantation. *N Engl J Med* 1991; **325**: 1601–7.

2 Goodrich JM, Bowden RA, Fisher L, Keller C, Schoch G, Meyers JD. Ganciclovir prophylaxis to prevent cytomegalovirus disease after allogeneic marrow transplant. *Ann Intern Med* 1993; **118**: 173–8.

3 Boeckh M, Huang M, Ferrenberg J, *et al.* Optimization of Quantitative Detection of Cytomegalovirus DNA in Plasma by Optimization of Quantitative Detection of Cytomegalovirus DNA in Plasma by Real-Time PCR. *J Clin Microbiol* 2004; **43**: 1142–8.

4 Therneau T. A Package for Survival Analysis in S, version 2.38. 2015. https://cran.r-project.org/package=survival%3E.

5 Gray B. cmprsk: Subdistribution Analysis of Competing Risks. R package version 2.2-7. https://cran.r-project.org/package=cmprsk.

6 R Core Team. R: A language and environment for statistical computing. 2018. https://www.r-project.org/.

7 Polley E, LeDell E, Kennedy C, Van Der Laan MJ; SuperLearner: Super Learner Prediction. R package version 2.0-24. 2018. https://cran.r-project.org/package=SuperLearner.

8 LeDell E, Petersen M, Van Der Laan MJ. cvAUC: Cross-Validated Area Under the ROC Curve Confidence Intervals. R package version 1.1.0. 2014. https://cran.r-project.org/package=cvAUC.

9 Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifer performance in R. *Bioinformatics* 2005; **21**: 7881.

10 Gelman A, Su Y-S. arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. 2018. https://cran.r-project.org/package=arm.

11 Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010; **33**: 1–22.

12 Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* 2015; **67**: 1–48.

13 Mullen KM, von Stokkum IHM. nnls: The Lawson-Hanson algorithm for non-negative least squares (NNLS). 2012. https://cran.r-project.org/package=nnls.

14 Chen T, He T, Benesty M, *et al.* xgboost: Extreme Gradient Boosting. 2019. https://cran.r-project.org/package=xgboost.

15 Grolemund G, Wickham H. Dates and Times Made Easy with lubridate. *J Stat Softw* 2011; **40**: 1–25.

16 Wickham H, Francois R. dplyr: A Grammar of Data Manipulation. R package version 0.5.0. https://cran.r-project.org/package=dplyr.

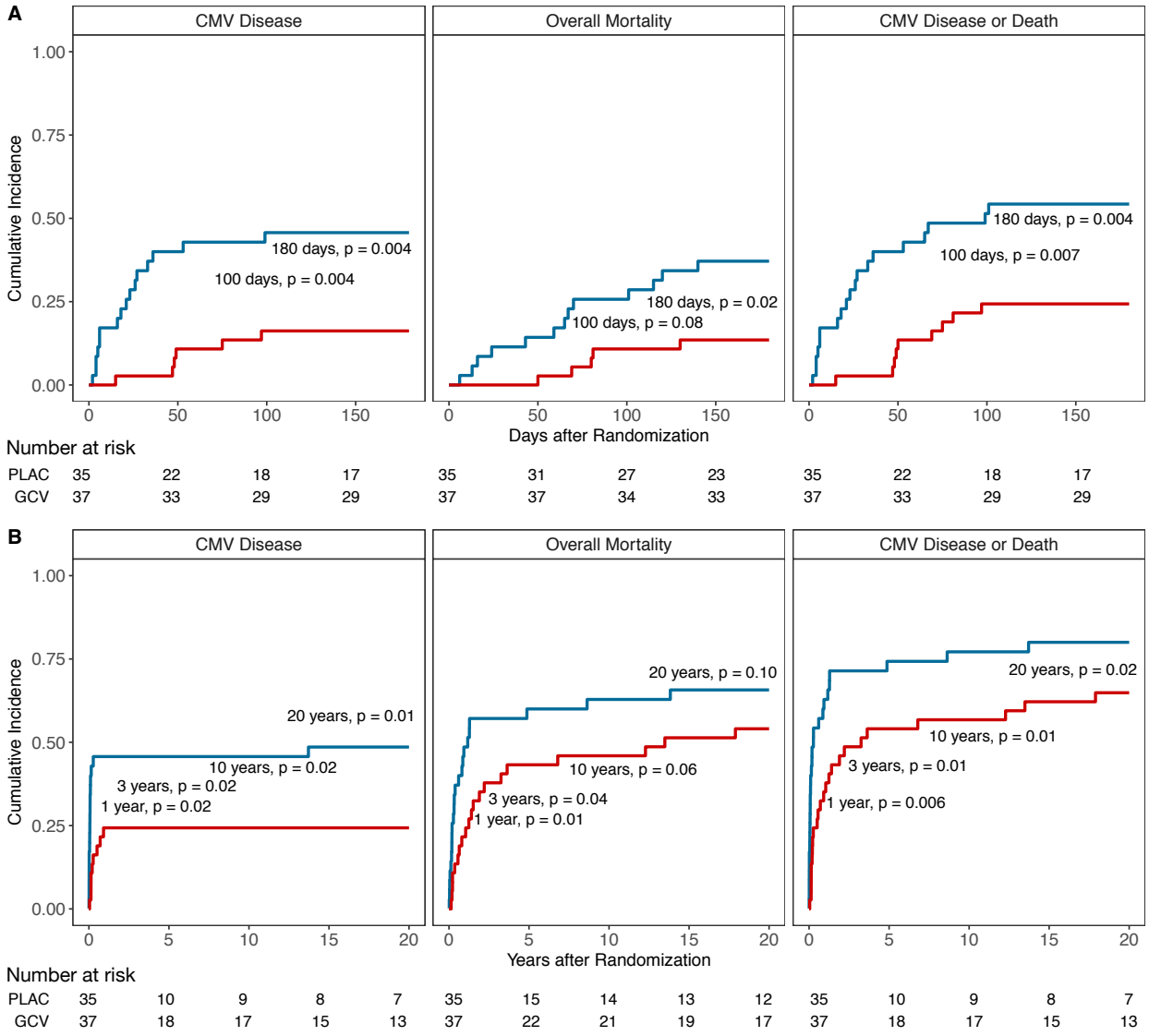
17 Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer Verlag, 2009.

18 Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. 2018. https://cran.r-project.org/package=cowplot.

19 Kobayashi F, Kuroki M. A new proportion measure of the treatment effect captured by candidate surrogate endpoints. *Stat Med* 2014; **33**: 3338–53.

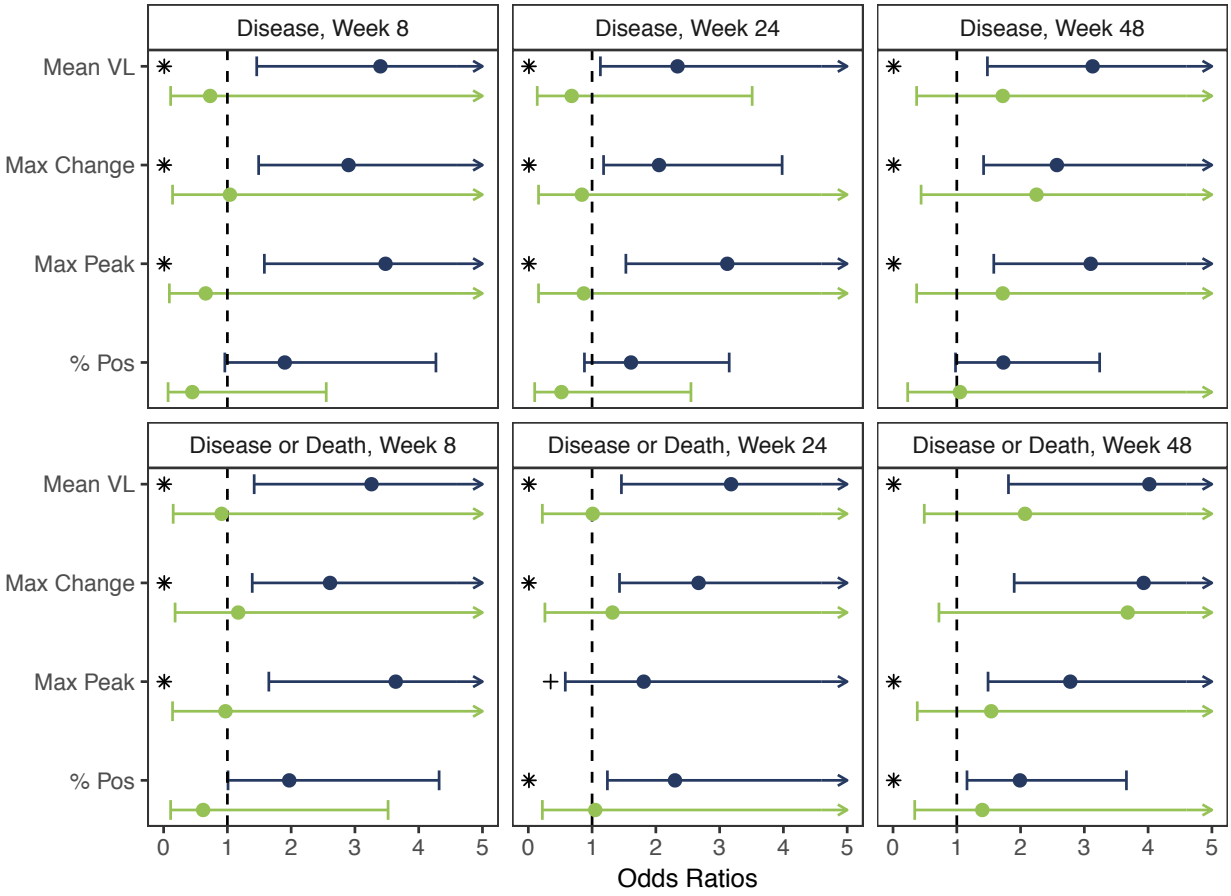
Supplementary Figures

Supplementary Figure 1



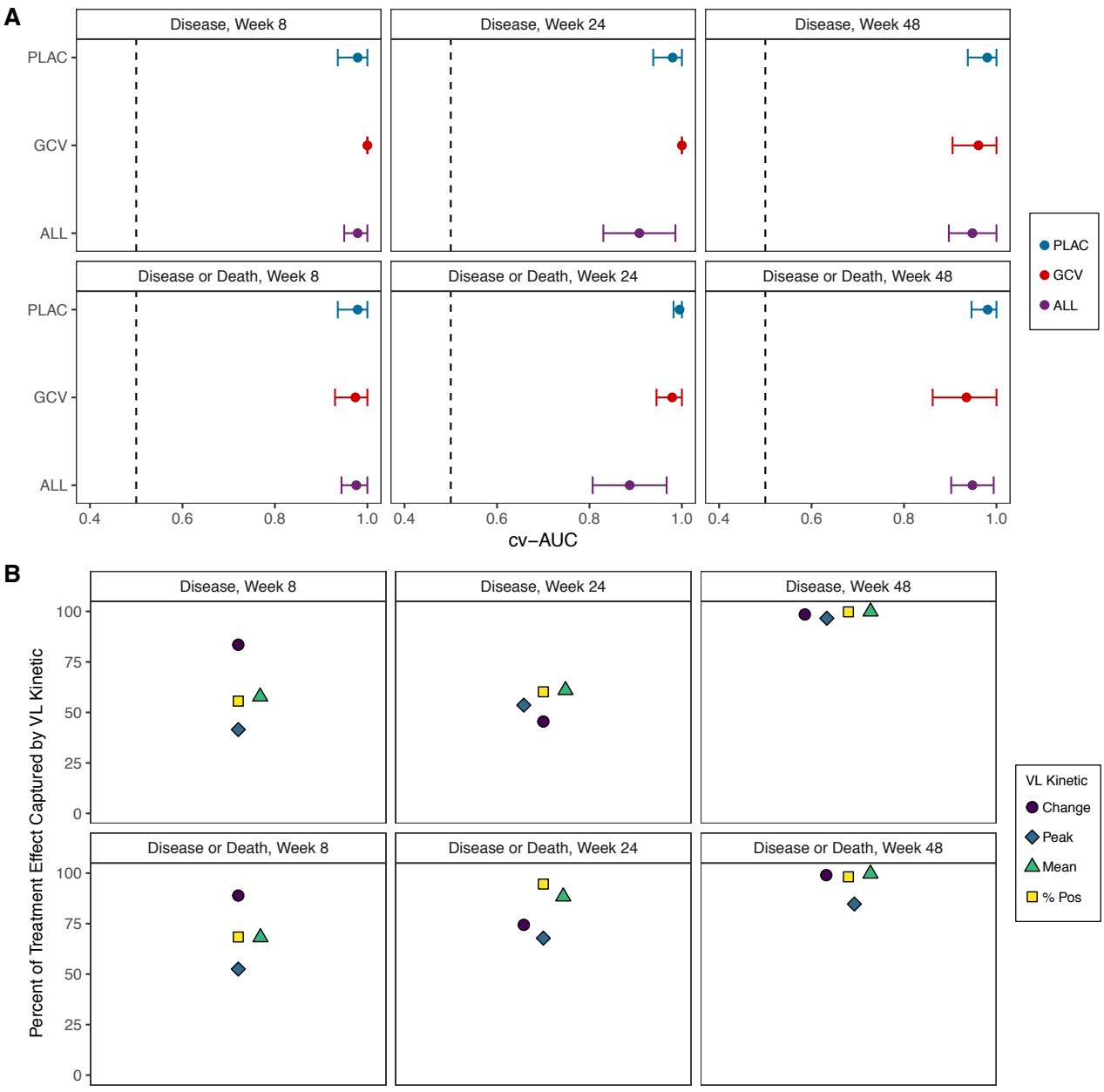
Supplementary Figure 1 – Cumulative incidence of CMV disease, overall mortality, and CMV disease or death by 180 days after randomization from NEJM 1991 RCT and by 20 years of extended follow up – (A) CMV disease, overall mortality, and the composite endpoint of first event of CMV disease or death in the placebo (blue) and ganciclovir (red) treatment groups by 100 and 180 days after randomization and (B) at extended follow up times out to 20 years after randomization. In (A & B) survival and first event of CMV disease or death curves were estimated using Kaplan-Meier methods. The cumulative incidence of CMV disease with death as a competing risk was estimated using the Aalen-Johnson method. Survival distributions and times to the composite endpoint of CMV disease or death were compared using the log-rank test. Cumulative incidence distributions for CMV disease with death as a competing risk were compared using Gray's test. Numbers at risk are shown below each plot.

Supplementary Figure 2



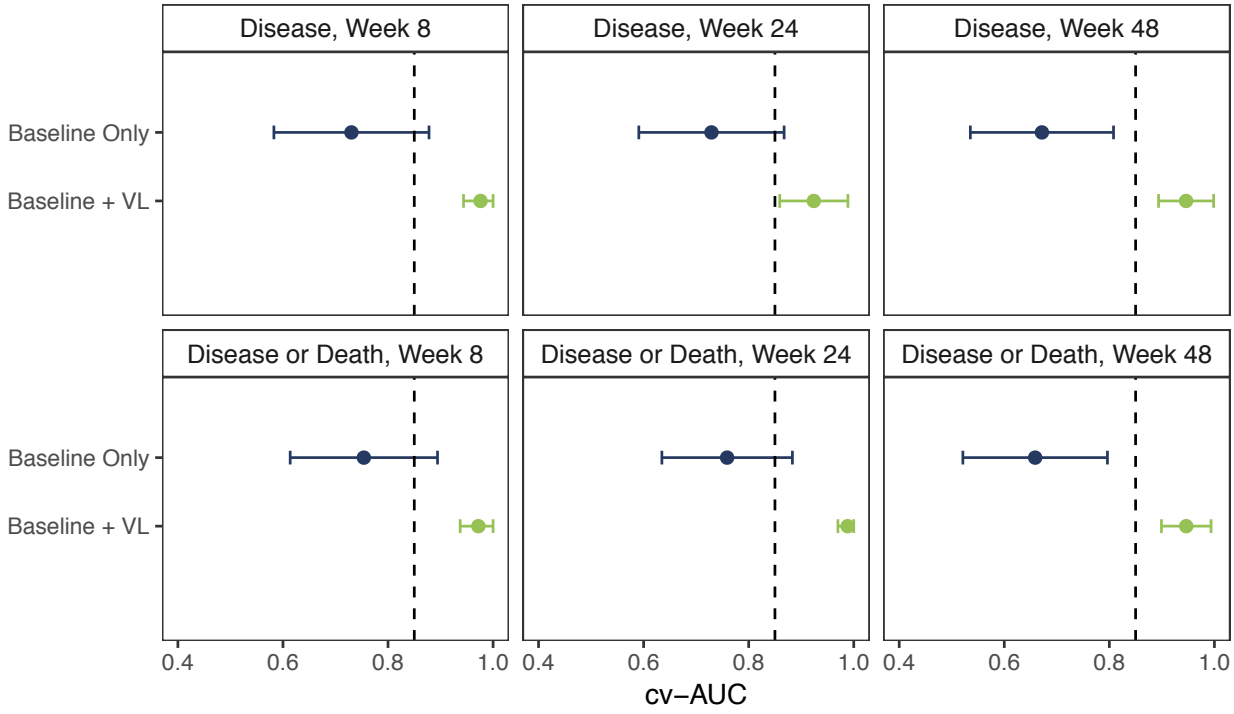
Supplementary Figure 2 - Prentice Criteria Evaluation Using Multivariate Logistic Regression. Forest plots of odds ratios (OR) for associations between VL kinetics and risk for CMV disease and CMV disease or death by weeks 8, 24, and 48 after treatment initiation calculated from logistic regression models adjusted for baseline characteristics and treatment group. OR for the VL kinetics are indicated by the navy dots surrounded by 95% confidence intervals (CI) indicated with navy lines; OR with 95% CI from treatment group assignment are in light green dots and lines. Asterisks (*) indicated VL kinetics in which the coefficient for the VL kinetic was significantly different from zero ($p < 0.05$), but the treatment group assignment coefficient was not significantly different from zero ($p \geq 0.20$) in multivariate logistic regression testing. For those VL kinetics that met these criteria, pluses (+) are shown when the interaction term coefficient was significantly different from zero ($p < 0.20$), indicating a potential interaction between the VL kinetic and treatment group. In this case, Prentice criteria are not met, and there is no common OR for the VL kinetic between treatment groups. Thus, only the OR for the ganciclovir group is shown. Mean VL is VL averaged over weeks 1-5; Max Change is the maximum difference between randomization VL and weeks 1-5 VL; Max Peak is the highest VL measured weeks 1-5; and % Pos is the percentage of VL measured weeks 1-5 with detectable VL. The vertical lines indicate OR = 1.

Supplementary Figure 3



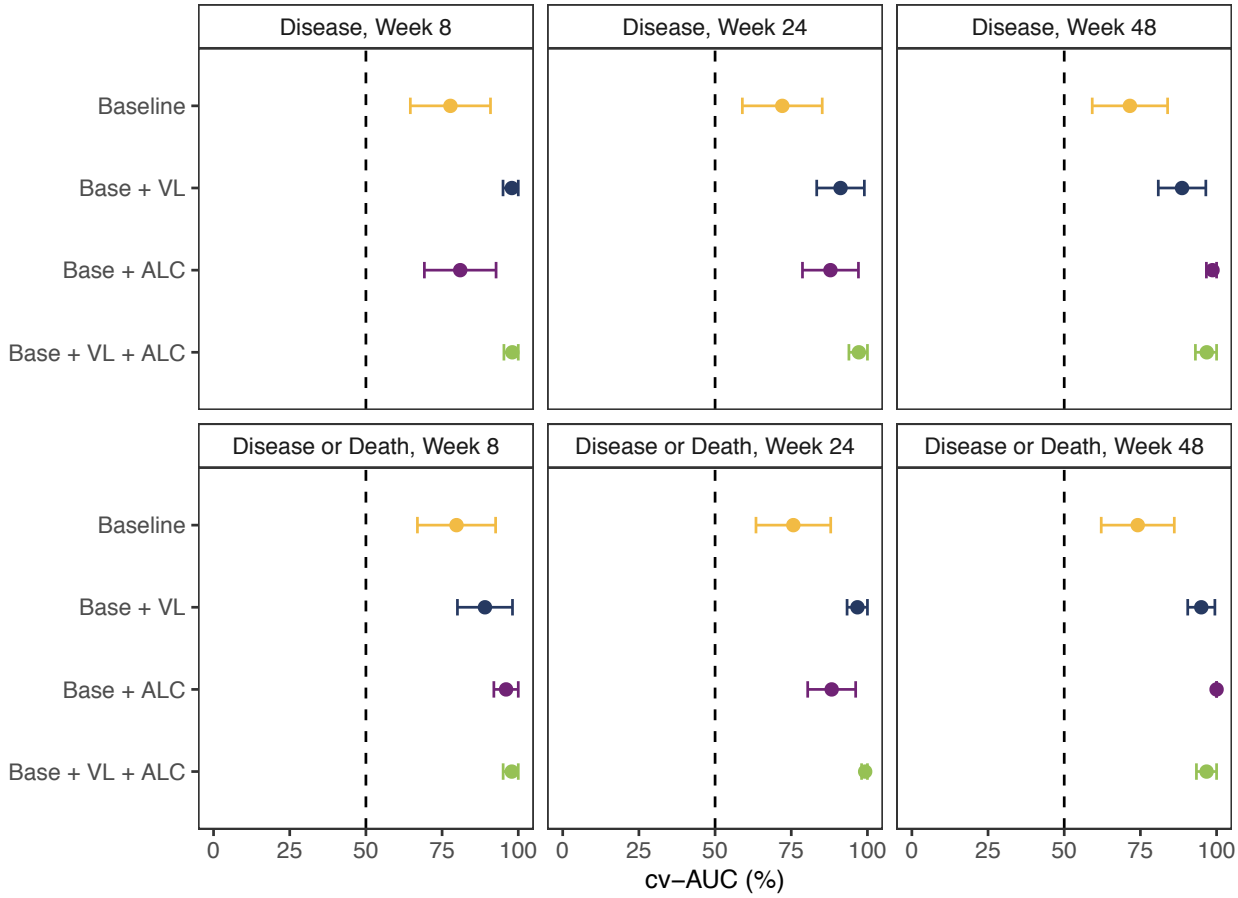
Supplementary Figure 3 – Prediction accuracy for clinical outcomes with Super Learners and percentage of ganciclovir effect captured. (A) Forest plots show cross-validated area under the receiver operator curves (cv-AUC) of Super Learner predictions for CMV disease and CMV disease or death. Predictions made only on data from the placebo group are in blue, from the ganciclovir group (GCV) in red, and from the treatment groups combined in purple. The vertical lines indicate cvAUC = 50%. (B) The percentage of ganciclovir's effect on clinical outcomes captured by the candidate surrogate is shown for each of the VL kinetics (shape and color).

Supplementary Figure 4



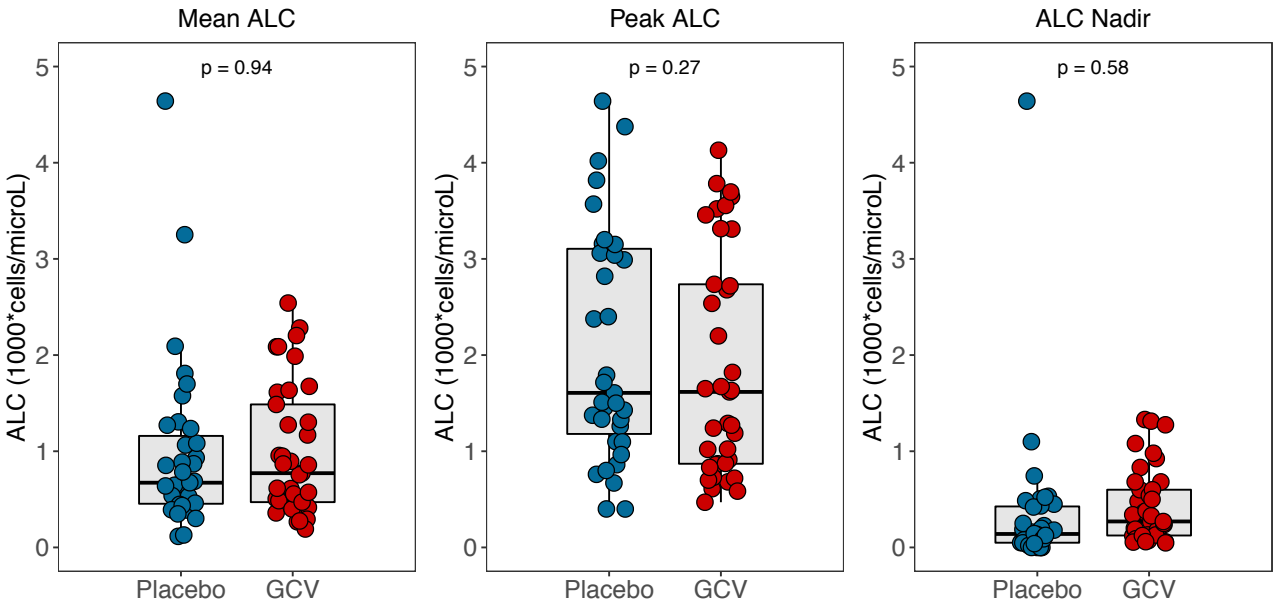
Supplementary Figure 4 – Prediction accuracy of Super Learners for clinical outcomes using baseline characteristics versus baseline characteristics plus all viral load kinetics. Forest plots compare cv-AUCs for Super Learner predictions using baseline data only in navy versus baseline data plus all VL kinetics (Mean, Change, Peak, % Positive) in light green. The vertical lines indicate cv-AUC = 85%.

Supplementary Figure 5



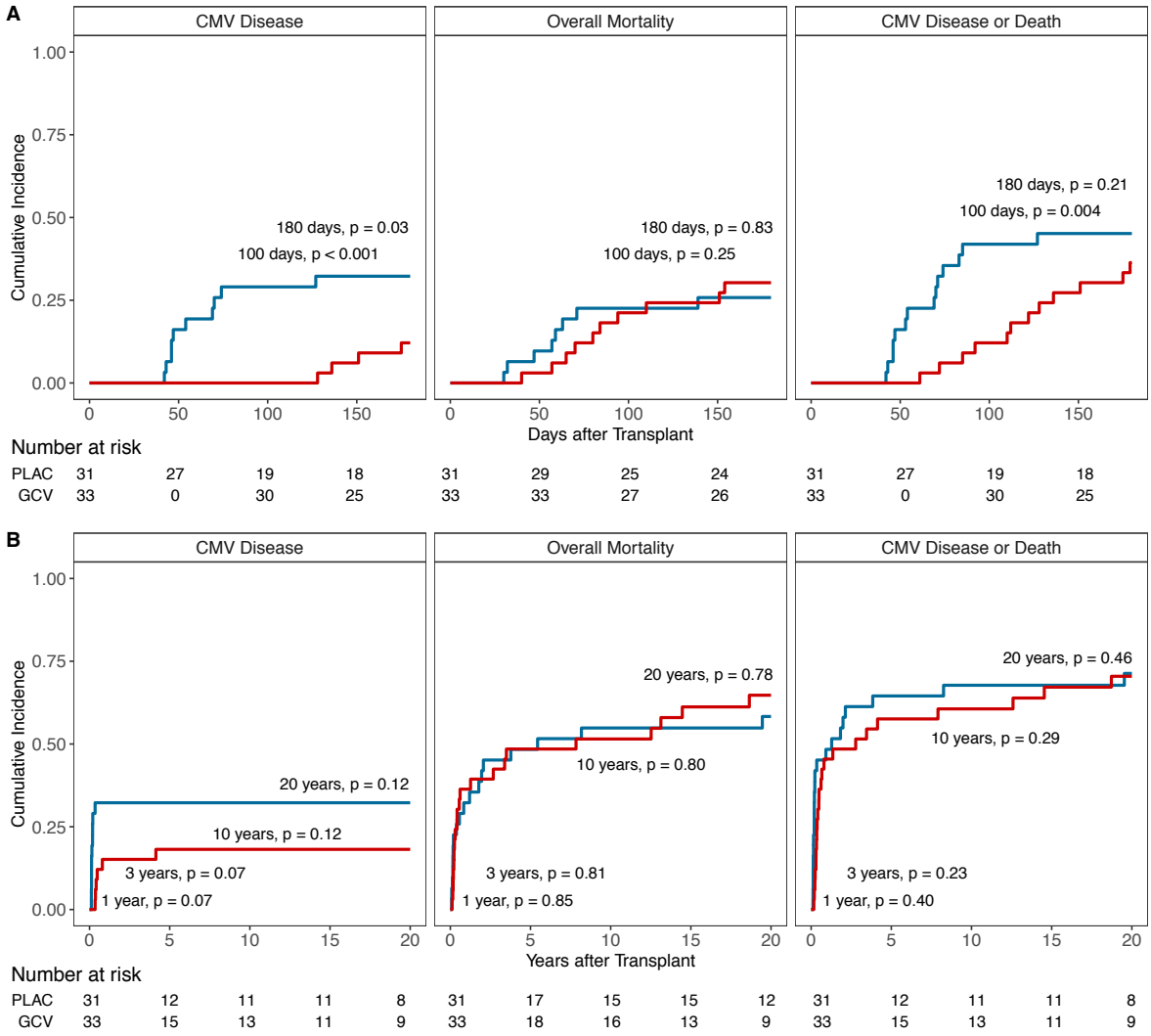
Supplementary Figure 5 – Prediction accuracy of Super Learners for clinical outcomes using baseline characteristics plus viral load kinetics plus absolute lymphocyte kinetics. Forest plots compare cv-AUCs for Super Learner predictions using baseline data only, "Base," (donor CMV serostatus, acute graft versus host disease, absolute lymphocyte count) in orange versus baseline data plus all VL kinetics (Mean, Change, Peak, % Positive), "Base + VL," in navy versus baseline data plus all absolute lymphocyte kinetics, "Base + ALC," versus the full model that includes baseline date plus all VL kinetics plus all ALC kinetics, "Base + VL + ALC." The vertical lines indicate cv-AUC = 50%.

Supplementary Figure 6



Supplementary Figure 6 – Absolute lymphocyte count (ALC) kinetics – ALC kinetics from time of randomization to five weeks post-randomization (and post-treatment). Box and whisker plots show the middle 50% of ALC kinetics in grey boxes with a horizontal black line at the median. Whiskers extend upward from the third quartile to 1.5 times the interquartile range (the distance between first and third quartiles) and downward from the first quartile to 1.5 times the interquartile range. p-values were calculated from t tests comparing the means of the ALC kinetics in ganciclovir (GCV) versus placebo groups.

Supplementary Figure 7



Supplementary Figure 7 – Cumulative incidence of CMV disease, overall mortality, and CMV disease or death by 180 days after randomization from AIM 1993 RCT and by 20 years of extended follow up – (A) CMV disease, overall mortality, and the composite endpoint of first event of CMV disease or death in the placebo (blue) and ganciclovir (red) treatment groups by 100 and 180 days after transplantation and (B) at extended follow up times out to 20 years after HCT. In (A & B) survival and first event of CMV disease or death curves were estimated using Kaplan-Meier methods. The cumulative incidence of CMV disease with death as a competing risk was estimated using the Aalen-Johnson method. Survival distributions and times to the composite endpoint of CMV disease or death were compared using the log-rank test. Cumulative incidence distributions for CMV disease with death as a competing risk were compared using Gray's test. Numbers at risk are shown below each plot.

Supplementary Tables

Supplementary Table 1 – Baseline characteristics of the early treatment study population stratified by treatment group

Characteristic	Ganciclovir	Placebo
No. of patients	37	35
Age – mean yr (range)	33 (2-56)	31 (3-51)
Sex – F/M*	20/17	15/20
Underlying disease – no. (%)		
Acute lymphocytic leukemia	4 (11)†	7 (20)
Acute nonlymphocytic leukemia	16 (43)	11 (31)
Chronic myelogenous leukemia	11 (30)	11 (31)
Hodgkin's disease	1 (3)	0 (0)
Non-Hodgkins lymphoma	2 (5)	4 (11)
Other	3 (8)	2 (6)
HLA matching – no. (%)		
Patient matched with related donor	21 (57)	28 (80)**
Patient matched with unrelated donor	5 (14)	4 (11)
Patient mismatched with donor	11 (30)	3 (9)
Acute GVHD – no. (%)‡		
Present	24 (65)§	24 (69)
Not present	13 (35)	11 (31)
CMV status before transplantation – no. (%)		
Patient negative, donor positive	3 (8)¶	3 (9)¥
Patient positive, donor negative	15 (41)	13 (37)
Patient and donor positive	19 (51)	19 (54)
Days from HCT to study entry – mean (range)	54 (18-79)	48 (16-77)

* In Goodrich et al. NEJM 1991[1], "M/F" should have been reported as "F/M."
† In ref[1], one additional patient is reported to have acute lymphocytic leukemia rather than "other."
** In ref[1], one fewer patient is reported as having a matched related donor; two additional patients are reported as having a matched unrelated donor; and one fewer is reported as having a mismatched donor.
‡ In Table 1, the presence of acute GVHD is counted from transplant to day 100. Between randomization and day 100, one patient in the ganciclovir group and one patient in the placebo group developed aGVHD.
§ In ref[1], one additional patient was reported to have aGVHD.
¶ In ref[1], three additional patients were counted as negative, donor positive; and three fewer patients are counted as positive with negative donors.
¥ In ref[1], one additional patient is reported as positive with negative donor; one fewer patient is reported as positive with positive donor.

Supplemental Table 2 – CMV disease or CMV disease or death events in patients with viral load data at weekly intervals following randomization

Event	Treatment Group	Number of Events		
Baseline (n = 72)		**Week 8**	**Week 24**	**Week 48**
CMV Disease	Placebo, n = 35	15	16	16
	GCV, n = 37	4	6	9
Disease/Death	Placebo, n = 35	15	19	21
	GCV, n = 37	5	9	13
Week 1 (n = 58)				
CMV Disease	Placebo, n = 24	10	11	11
	GCV, n = 34	4	6	9
Disease/Death	Placebo, n = 24	10	13	14
	GCV, n = 34	5	9	13
Week 2 (n = 56)				
CMV Disease	Placebo, n = 25	9	10	10
	GCV, n = 31	4	4	7
Disease/Death	Placebo, n = 25	9	11	13
	GCV, n = 31	5	7	11
Week 3 (n = 55)				
CMV Disease	Placebo, n = 21	6	7	7
	GCV, n = 34	3	5	8
Disease/Death	Placebo, n = 21	6	8	10
	GCV, n = 34	4	8	12
Week 4 (n = 46)				
CMV Disease	Placebo, n = 18	4	5	5
	GCV, n = 28	2	4	6
Disease/Death	Placebo, n = 18	4	5	7
	GCV, n = 28	2	5	8
Week 5 (n = 37)				
CMV Disease	Placebo, n = 15	3	3	3
	GCV, n = 22	2	3	5
Disease/Death	Placebo, n = 15	3	4	6
	GCV, n = 22	2	3	5
Summary Kinetics (n = 65)				
CMV Disease	Placebo, n = 29	10	11	11
	GCV, n = 36	4	6	9
Disease/Death	Placebo, n = 29	10	13	15
	GCV, n = 36	5	9	13

Supplemental Table 3 – Risk for CMV disease or death by multivariable logistic regression adjusted for aGVHD, CMV donor serostatus, and randomization viral load in the early treatment trial

CMV Disease by Week 8

VL Kinetic	OR	95% CI	p-value
Mean VL	3.7	(1.8, 9.2)	0.001
Peak	3.8	(1.9, 9.9)	0.001
Max Change	2.9	(1.7, 5.8)	<0.001
Percent Pos	2.3	(1.3, 4.7)	0.009

CMV Disease or Death by Week 8

VL Kinetic	OR	95% CI	p-value
Mean VL	3.3	(1.7, 7.9)	0.002
Peak	3.7	(1.9, 9.3)	0.001
Max Change	2.5	(1.5, 4.8)	0.001
Percent Pos	2.2	(1.3, 4.4)	0.009

CMV Disease by Week 24

VL Kinetic	OR	95% CI	p-value
Mean VL	2.6	(1.4, 5.4)	0.004
Peak	3.2	(1.7, 7.3)	0.001
Max Change	2.1	(1.4, 3.7)	0.002
Percent Pos	1.9	(1.2, 3.3)	0.020

CMV Disease or Death by Week 24

VL Kinetic	OR	95% CI	p-value
Mean VL	3.2	(1.6, 7.3)	0.002
Peak	3.0	(1.6, 6.5)	0.001
Max Change	2.5	(1.5, 4.7)	0.002
Percent Pos	2.3	(1.4, 4.1)	0.003

CMV Disease by Week 48

VL Kinetic	OR	95% CI	p-value
Mean VL	2.7	(1.5, 5.5)	0.003
Peak	2.7	(1.6, 5.5)	0.002
Max Change	2.1	(1.4, 3.6)	0.002
Percent Pos	1.7	(1.1, 2.8)	0.030

CMV Disease or Death by Week 48

VL Kinetic	OR	95% CI	p-value
Mean VL	3.2	(1.7, 7.0)	0.001
Peak	2.5	(1.5, 4.7)	0.002
Max Change	2.7	(1.6, 5.1)	<0.001
Percent Pos	1.8	(1.2, 3.0)	0.007

Supplementary Table 4 – Baseline characteristics of the prophylaxis study population stratified by treatment group

CHARACTERISTIC	GANCICLOVIR	PLACEBO
No. of patients	33	31
Age – mean in years (range)	34 (4-55)	34 (3-56)
Sex - *n*		
Male	17	16
Female	16	15
Underlying disease – *n*		
Acute lymphocytic leukemia	2*	2
Acute myelogenous leukemia	7	11
Myelodysplastic syndrome	3	3
Chronic myelogenous leukemia	16	13
Lymphoma	4*	1
Multiple Myeloma	1	1
HLA matching – *n*		
Patient matched with related donor	16	17
Patient matched with unrelated donor	10	6
Patient mismatched with donor	7	8
Acute GVHD – *n*		
Grades 0 - 1	14	12
Grades 2 - 4	19	19†
CMV donor serostatus – *n*		
Positive	15	15
Negative	18	16
Days from transplantation to study entry – mean (range)	25 (15-38) ‡	24 (13-35)

* In Goodrich et al. AIM 1993[2], the ganciclovir group was reported to include 1 patient with ALL and 5 patients with lymphoma
† In ref[2], the placebo group is listed as having 19 patients with grades 0-1 and 19 patients with grades 2-4 graft-versus-host-disease, but should be 12 and 19, respectively as above.
‡ In ref[2], the ganciclovir group was reported to have included patients randomized from 2 to 38 days, which should instead be 15 to 38 days as above.

Supplemental Table 5 – Risk for CMV disease or death by multivariable logistic regression adjusted for aGVHD, CMV donor serostatus, and randomization viral load in the prophylaxis trial

CMV Disease by Week 24

Viral Load Kinetic	Odds Ratio (95% CI)	p-value
Mean	2.6 (1.2, 6.3)	0.02
Peak	1.8 (1.2, 3.0)	0.01
Maximum Change	1.8 (1.2, 3.0)	0.01
Percent Positive	1.8 (1.2, 2.9)	0.007