JCI The Journal of Clinical Investigation

Cytomegalovirus viral load kinetics as surrogate endpoints after allogeneic transplantation

Elizabeth R. Duke, … , Joshua T. Schiffer, Michael Boeckh

J Clin Invest. 2020. https://doi.org/10.1172/JCI133960.

Clinical Medicine In-Press Preview **Clinical trials** **Infectious disease**

Background: Viral load surrogate endpoints transformed development of HIV and hepatitis C therapeutics. Surrogate endpoints for cytomegalovirus (CMV)-related morbidity and mortality could advance development of antiviral treatments. While observational data support using CMV viral load (VL) as a trial endpoint, randomized controlled trials (RCT) demonstrating direct associations between virologic markers and clinical endpoints are lacking.

Methods: We performed CMV DNA polymerase chain reaction (PCR) on frozen serum samples from the only placebo-controlled RCT of ganciclovir for early treatment of CMV after hematopoietic cell transplantation (HCT). We used established criteria to assess VL kinetics as surrogates for CMV disease or death by weeks 8, 24, and 48 after randomization and quantified antiviral effects captured by each marker. We used ensemble-based machine learning to assess the predictive ability of VL kinetics and performed this analysis on a ganciclovir prophylaxis RCT for validation.

Results: VL suppression with ganciclovir reduced cumulative incidence of CMV disease and death for 20 years after HCT. Mean VL, peak VL, and change in VL during the first five weeks of treatment fulfilled the Prentice definition for surrogacy, capturing > 95% of ganciclovir's effect, and yielded highly sensitive and […]

Cytomegalovirus viral load kinetics as surrogate endpoints after allogeneic transplantation

Elizabeth R Duke, MD[1,2], Brian D Williamson, PhD[1], Bhavesh Borate, MS[1], Jonathan L Golob, MD, PhD[1,2], Chiara Wychera[1], Terry Stevens-Ayers, MS[1], Meei-Li Huang, PhD[2], Nicole Cossrow, PhD[3], Hong Wan, PhD[3], T Christopher Mast, PhD[3], Morgan A Marks, PhD[3], Mary E Flowers, MD[1,2], Keith R Jerome, MD, PhD[1,2], Lawrence Corey, MD[1,2], Peter B Gilbert, PhD[1,2], Joshua T Schiffer, MD[1,2*], Michael Boeckh, MD, PhD[1,2‡*]

1. Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA, 98109, USA
2. University of Washington, 1959 NE Pacific St, Seattle, WA, 98195, USA
3. Merck & Co., Inc., 2000 Galloping Hill Rd, Kenilworth, NJ, 07033, USA

Jonathan Golob's current affiliation is University of Michigan, 500 S State St, Ann Arbor, MI, 48109, USA.

*Joshua Schiffer and Michael Boeckh contributed equally.

‡Corresponding Author:
Michael Boeckh, MD, PhD
1100 Fairview Ave N, E4-100
Seattle, WA 98109
mboeckh@fredhutch.org
206-667-4898

Conflict of Interest Statement

The Fred Hutch (ERD, BB, TSA, MEF, KRJ, LC, PBG, JTS, MJB) received a research grant from Merck & Co., Inc. to support the conduct of the study. NC, HW, TCM, MAM are employees of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA. BDW, BB, JLG declare no competing interests. Conflicts of interest not related to the submitted study include KRJ: grant funding from Immune Design and personal fees from Editas Medicine, Inc, Janssen Sciences Ireland, and Vir Biotechnology; LC: ownership interest in Immune Design, Vir Biotechnology, Inc., and Cero Therapeutics; PGB: grant funding from Sanofi Pasteur; JTS reports grant funding from Genocea Biosciences, Inc.; MB: grant funding from Astellas, Chimerix, Gilead, GlaxoSmithKline, Takeda, VirBio, and Lophius Biosciences, consultant to Allovir, Artemis Therapeutics, Astellas, Chimerix, Gilead, GlaxoSmithKline, Moderna Therapeutics, Oxford Immunotec, Takeda, and VirBio, and scientific advisory board member of Helocyte and Evrys Bio.

Abstract

Background:

Viral load surrogate endpoints transformed development of HIV and hepatitis C therapeutics. Surrogate endpoints for cytomegalovirus (CMV)-related morbidity and mortality could advance development of antiviral treatments. While observational data support using CMV viral load (VL) as a trial endpoint, randomized controlled trials (RCT) demonstrating direct associations between virologic markers and clinical endpoints are lacking.

Methods:

We performed CMV DNA polymerase chain reaction (PCR) on frozen serum samples from the only placebo-controlled RCT of ganciclovir for early treatment of CMV after hematopoietic cell transplantation (HCT). We used established criteria to assess VL kinetics as surrogates for CMV disease or death by weeks 8, 24, and 48 after randomization and quantified antiviral effects captured by each marker. We used ensemble-based machine learning to assess the predictive ability of VL kinetics and performed this analysis on a ganciclovir prophylaxis RCT for validation.

Results:

VL suppression with ganciclovir reduced cumulative incidence of CMV disease and death for 20 years after HCT. Mean VL, peak VL, and change in VL during the first five weeks of treatment fulfilled the Prentice definition for surrogacy, capturing > 95% of ganciclovir's effect, and yielded highly sensitive and specific predictions by week 48. In the prophylaxis trial, viral shedding rate satisfied the Prentice definition for CMV disease by week 24.

Conclusion:

Our results support using CMV VL kinetics as surrogates for CMV disease, provide a framework for developing CMV preventative and therapeutic agents, and support reductions in viral load as the mechanism through which antivirals reduce CMV disease.

Funding:

Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA, National Institutes of Health, and Fred Hutch Vaccine and Infectious Disease Division

Introduction

Despite advances in the treatment and prevention of cytomegalovirus (CMV) complications after HCT, CMV remains an important cause of morbidity and mortality. CMV viremia is associated with increased non-relapse mortality[1,2], acute graft versus host disease (aGVHD)[3], and secondary bacterial and fungal infections[4,5]. Since the 1990s, HCT recipients have been treated pre-emptively with antiviral drugs for the prevention of CMV disease. However, all antivirals approved for the treatment of CMV disease and for pre-emptive therapy (ganciclovir, valganciclovir, and foscarnet) cause significant toxicities, including neutropenia, renal failure, and genital ulcers)[6,7]. Additional antiviral therapies are needed to reduce CMV-related complications following HCT[6]. Establishing CMV viral load-based surrogate endpoints for use in clinical trials would facilitate development of new antiviral therapeutics[7,8]. Indeed, the well-tolerated, antiviral drug letermovir was recently approved for preventing CMV reactivation as prophylactic therapy using a combined endpoint that included clinically significant infection (CMV viral load at a level high enough to warrant pre-emptive therapy)[9].

Surrogate endpoints are biomarkers that predict clinical outcomes accurately enough to replace those outcomes in clinical trials. Using surrogate endpoints in clinical trials can reduce follow-up time and the number of patients required to demonstrate an effect, reducing research costs and burden to trial participants and facilitating delivery of new therapies to bedside[10,11]. FDA approval of VL surrogate endpoints revolutionized antiviral drug development for HIV and hepatitis C[8,10]. Clinical trials for HIV and hepatitis C now use viral load-based endpoints, a practice that has dramatically reduced times to licensure of new antivirals[12–14].

HIV and hepatitis C VL surrogates were validated via meta-analyses of large numbers of RCTs that were performed during the era of VL testing with PCR[15,16]. A recent meta-analysis in solid organ transplantation (SOT) has provided evidence that viral load may be a valid surrogate for CMV disease in the SOT setting, but lacks placebo-controlled RCT data[8]. In the HCT setting, prior to our study, no VL data from placebo-controlled, randomized treatment trials existed, as

these trials were conducted long before the availability of PCR testing[17–21]. Validating viral load-based surrogates for CMV is not possible in the modern clinical environment both due to the absence of placebo-controlled antiviral trials (for equipoise, ganciclovir and foscarnet are used as active controls based on their proven association with clinical benefit) and due to small numbers of clinical CMV disease cases. However, despite changes in HCT care, CMV viral reactivation continues to occur in the modern setting and likely remains the primary mechanism through which CMV disease occurs[1,2]. To address whether viral load-based surrogates are valid surrogate endpoints, we performed VL testing of frozen samples obtained during a historic clinical trial—the first and only double-blind, placebo-controlled, randomized trial for the early treatment of CMV infection with ganciclovir after bone marrow transplantation—and calculated CMV VL kinetics to assess their potential use as surrogate endpoints[17].

We employed traditional statistical methods and state-of-the-art machine learning techniques to validate viral load as a surrogate endpoint. The Prentice definition (traditional methodology) is a rigorous statistical standard for evaluating whether an intermediate response endpoint is a valid surrogate endpoint[11,22,23]. We applied the Prentice definition to our data to evaluate whether VL kinetics could serve as valid surrogate endpoints and quantified the degree to which they captured ganciclovir's effect on clinical outcomes. In addition, we employed ensemble-based machine learning models (Super Learners[24]) to determine the ability of viral load kinetics to predict clinical outcomes. Finally, given that many centers now use prophylaxis as their primary CMV prevention strategy, we used these same techniques to validate our results in the prophylactic setting with patient samples from the first ganciclovir prophylaxis RCT[18].

Results

Ganciclovir reduced CMV disease and mortality at least 20 years after the original RCT

In a single-center study performed at the Fred Hutchinson Cancer Research Center (Fred Hutch) from 1989 to 1990,[17] seventy-two allogeneic HCT recipients who were either CMV seropositive or who had received marrow from CMV seropositive donors were screened weekly for CMV with viral cultures and were randomized to receive either ganciclovir or placebo at the time of first positive culture. A description of the study design is provided in Figures 1A-B, and baseline patient characteristics are shown in Supplemental Table 1. A schematic of the viral load analysis is shown in Figure 1C.

The original trial was designed to enroll 116 patients but was terminated early after the interim analysis showed a large reduction in tissue-invasive CMV disease by 100 days after HCT. Ganciclovir was found to have reduced significantly the cumulative incidence of CMV disease and overall mortality at 100 and 180 days after HCT (Figure 2A). Extending follow up of results observed in the original RCT through chart review, we found that the cumulative incidence of CMV disease and of the composite endpoint of CMV disease or death remained significantly lower in the ganciclovir group after 20 years (Figure 2B). Overall mortality was also lower in the ganciclovir group after 20 years (Figure 2B) though the trend in mortality was no longer statistically significant by 10 years. When outcomes were counted from randomization rather than transplantation, results were similar (Supplemental Figure S1). Detailed methods and results from the original study and extended follow up are included in the Supplemental Text. By providing evidence of a successful intervention in an RCT, these results demonstrate that the Prentice definition can be applied to our data.

Ganciclovir lowered CMV viral load kinetics in the first five weeks after randomization

Validation of surrogate endpoints requires the measurement of candidate biomarkers at intermediate time points after randomization. Frozen serum samples leftover from clinical testing were stored prospectively in a biorepository at Fred Hutch for all study participants. CMV DNA PCR viral loads were measured from available samples collected at approximately weekly

intervals up to day 100 after HCT. Viral loads collected near the time of randomization until the first event of CMV disease were included in the surrogate analysis. All 72 patients had viral load samples available near the time of randomization. Sixty-five patients had at least one viral load measured in weeks 1 through 5 with a median of 4 measurements per patient in both treatment groups. Detailed sample availability information is provided in Supplemental Table 2.

CMV viral load kinetics, including mean viral load (log 10 IU/mL), maximum change in viral load from randomization (log 10 IU/mL), peak viral load (log 10 IU/mL), and percentage of positive viral loads (viral shedding rate) were calculated from viral loads measured in the first five weeks following randomization. Only early viral loads (weeks 1 – 5) were included because surrogate endpoints are more useful when measured early after interventions and because many patients in the placebo group died or developed CMV disease soon after randomization. Weekly mean viral loads and changes in viral load, and all viral load kinetics are shown in figure 3.

CMV viral load kinetics fulfill the Prentice definition for valid surrogate endpoints

We evaluated whether each of the four viral load kinetics defined above (mean, maximum change, peak, and percent positive) is a valid surrogate for the clinical endpoints of tissue-invasive CMV disease or the composite endpoint of CMV disease or death by 8, 24, and 48 weeks after randomization (Figure 1C). In the manuscript text, we focus on results for week 48. Results for weeks 8 and 24 are provided in the Supplement. Because patients were randomized at varying times from HCT based on positive viral culture results, we chose weeks rather than days to describe outcomes in the surrogate analysis to help differentiate time from randomization rather than time from transplantation.

We validated each viral load kinetic based on fulfillment of the Prentice definition for valid surrogate endpoints. The Prentice definition requires that a hypothesis test of the treatment effect (e.g. ganciclovir effect) on the surrogate endpoint (e.g. viral load kinetic) is a

valid hypothesis test of the treatment effect on the clinical endpoint (e.g. CMV disease). In other words, if a clinical trial assessing the effect of a treatment was performed with the primary outcome being an effect on the surrogate marker rather than the clinical outcome, the overall conclusion would be the same as a study performed using the clinical endpoint[22,23].

Prentice criterion 1

To satisfy the Prentice definition, surrogates must fulfill three main criteria. The first criterion requires that treatment (ganciclovir) impacts the candidate surrogate endpoint (e.g. peak viral load). This criterion was met for all viral load kinetics as reported above and in Figure 3C, as mean, maximum change, peak, and percentage of positive viral loads were significantly lower in the ganciclovir group.

Prentice criterion 2

The second Prentice criterion is met if there is an association between candidate surrogates (viral load kinetics) and clinical outcomes (CMV disease or death). Logistic regression models adjusted for aGVHD, CMV donor serostatus, and viral load at randomization, but not for treatment group assignment, demonstrated that all viral load kinetics met this criterion for all clinical endpoints at weeks 8, 24, and 48 (Supplemental Table 3), i.e. higher values of the viral load kinetics correlated with significantly higher odds of CMV disease or death.

Prentice criterion 3

The third Prentice criterion states that for a given value of the candidate surrogate (e.g. maximum change in viral load), the probability of the clinical outcome (e.g. CMV disease) is the same in each treatment group (ganciclovir or placebo group). We tested for fulfillment of this criterion with logistic regression models adjusted for aGVHD, CMV donor serostatus, viral load at randomization, and treatment group. Because we adjusted for treatment group assignment

and viral load kinetics in these models, we were able to determine whether the treatment group assignment correlated with outcomes *after adjustment for the viral load kinetic*. Thus, to fulfill Prentice criterion 3, the odds ratio (OR) for the viral load kinetic should be significantly greater than one at the p = 0.05 level, and the odds ratio for the treatment assignment should not differ significantly from one at the p = 0.2 level (p-value threshold higher to demonstrate similarity in values rather than difference). Figure 4 illustrates with asterisks that mean viral load, peak viral load, and maximum change in viral load met Prentice criterion 3 ($p < 0.05$ for VL association; $p \geq 0.20$ for treatment group association) for CMV disease by week 48 with no evidence of a treatment by marker interaction ($p \geq 0.20$). Percentage of positive viral loads nearly satisfied Prentice criteria (p = 0.07 for VL association). Mean, peak, and percentage of positive viral loads also satisfied Prentice criteria for the composite outcome by week 48. Maximum change in viral load did not meet Prentice criterion 3 for the composite outcome, as p = 0.14 for treatment group association. Results for clinical endpoints occurring by weeks 8 and 24 were similar and are shown in Supplemental Figure 2.

Viral load kinetics capture a large proportion of ganciclovir's effect on clinical outcomes

We quantified how much of ganciclovir's effect on clinical outcomes could be attributed to its effects on viral load kinetics using the proportion of treatment effect captured by candidate surrogate endpoints[23]. For the week 48 clinical outcome of CMV disease, several viral load kinetics captured nearly all of the effect of ganciclovir: mean (99.9%), change (96.6%), peak (98.5%), and percent positive (95.8%) (Figure 4B). Mean, maximum change, and percent positive captured at least 93% of ganciclovir's effect on the composite outcome of CMV disease or death at week 48, whereas peak captured 84.5% (Figure 4B). Almost all viral load kinetics were considered "moderate" (> 63%) or "substantial" (> 85%) for composite outcomes by weeks 8 and 24. Maximum change captured 83.5% of ganciclovir's effect on CMV disease by week 8,

256 but other kinetics did not perform well for CMV disease by weeks 8 and 24 (Supplemental
257 Figure 3B).
258
259 *Super Learners predict clinical outcomes with high accuracy*
260 The Super Learner is a cross-validation-based ensemble machine learning method for
261 estimating the optimal weighted average of the predictions from a library of algorithms. Each of
262 these algorithms estimates the conditional probability of an event (e.g. CMV disease or no CMV
263 disease) given a set of potential risk factors (e.g. viral load kinetics or baseline risk factors)
264 using cross-validation[24,25]. For surrogate validation, in addition to providing optimal prediction
265 accuracy, Super Learner predictions have the advantage of evaluating the ability of surrogate
266 endpoints to predict clinical outcomes for individuals, rather than describing mean behavior on
267 the population level[26]. We built Super Learner models using baseline covariates (aGVDH, CMV
268 donor serostatus, and viral load at randomization) and all viral load kinetics (mean, maximum
269 change, peak, percent positive). As an exploratory analysis, we also fit Super Learners using
270 absolute lymphocyte kinetics.
271 We constructed receiver operating characteristic curves (ROCs) to evaluate the
272 sensitivity and specificity of Super Learner predictions for clinical outcomes and assessed their
273 performance with leave-one-out cross-validated area under the ROCs (cv-AUC). cv-AUC can be
274 interpreted as the probability that a randomly-selected patient experiencing a clinical outcome
275 will have a higher predicted risk than a randomly-selected patient not experiencing the outcome.
276 Models that predict at the same level of accuracy as random chance have cv-AUC = 50%.
277 Super Learners predicting both week 48 clinical outcomes yielded cv-AUC > 90% (Figure 5A-
278 D). All models built on mean, maximum change, peak, and percent positive viral loads, whether
279 fit separately on treatment groups or on the combined data set, predicted both clinical outcomes
280 (CMV disease/CMV disease or death) at all time points (weeks 8, 24, and 48) with better than
281 85% cv-AUC (Supplemental Figure 3A). Our results suggest that viral load kinetics measured

during the first 5 weeks of antiviral treatment combined with an ensemble machine learning algorithm allow for excellent clinical outcome prediction. In addition, models built on the placebo, ganciclovir, and combined groups performed similarly, consistent with the Prentice definition.

To evaluate the contributions of VL kinetics to the accuracy of the Super Learner predictions, we fit Super Learners using baseline characteristics only versus baseline characteristics plus all VL kinetics. We found that adding all VL kinetics to the baseline characteristics increased prediction accuracy greatly for all time points and both clinical outcomes (Supplemental Figure 4). For example, the model built on baseline characteristics alone had a cv-AUC of 75.5% (95% CI: 61-90%) for CMV disease or death by week 8, but the cv-AUC increased to 96.8% (95% CI: 93-100%) when viral load kinetics were included.

Including absolute lymphocyte counts in the Super Learners improves prediction of some clinical outcomes

We calculated absolute lymphocyte count (ALC) kinetics, including ALC peak, ALC nadir, and mean ALC, during the five-week period after randomization to explore whether adding longitudinal measures of immunity to the machine learning models might improve prediction accuracy for clinical outcomes[27]. In addition, we added ALC at randomization to the baseline risk characteristics (donor CMV serostatus, aGVHD, and viral load at randomization). We found that adding ALC kinetics did not change prediction accuracy of CMV disease by earlier time points (weeks 8 and 24), but improved prediction of CMV disease by week 48. ALC also improved prediction of CMV disease or death at all time points (Supplemental Figure 5). However, importantly, absolute lymphocyte kinetics did not consistently increase or decrease with ganciclovir administration (Supplemental Figure 6), and thus cannot be assessed as a surrogate for antiviral treatment. A surrogate of treatment effect must be affected in a consistent direction by the intervention.

Validation analysis performed from the ganciclovir prophylaxis RCT demonstrates viral load kinetics are valid surrogates in the prophylaxis setting

As follow-up to the early treatment trial, ganciclovir was studied as a prophylactic agent in a placebo-controlled RCT at the Fred Hutch from 1990 to 1991. Sixty-four CMV seropositive allogeneic HCT recipients were randomized to receive ganciclovir or placebo at engraftment and were followed for development of CMV infection (by culture) and CMV disease[18]. The CONSORT diagram and trial design schematic are shown in Figures 6A & B. Baseline patient characteristics are shown in Supplemental Table 4. We analyzed clinical outcomes by weeks 14, 24, and 48. The cumulative incidence of CMV disease was significantly lower in the ganciclovir treatment group by weeks 14 and 24, but no difference in mortality was found at these or later times points (Figure 7). The same results are shown in Supplemental Figure 7 in days and years from transplant rather than randomization. The same viral load kinetics: mean, peak, maximum change, and percentage of positive viral loads (shedding rate) were calculated for the first five weeks following randomization. As in the early treatment RCT, all viral load kinetics were significantly lower in the ganciclovir group, fulfilling Prentice criterion 1 (Figure 8).

Because no CMV disease events occurred in the treatment group during the first 14 weeks of the study, we were unable to perform the analyses at this time point. Thus, CMV disease by week 24 served as our primary clinical outcome. Prentice criterion 2 was met for all viral load kinetics by week 24 (Supplemental Table 5). Only the percentage of positive viral loads (shedding rate) met Prentice criterion 3, demonstrating a significant association between viral load after adjustment for treatment group (Figure 9A). However, the remaining viral load kinetics: mean, peak, and maximum change, nearly fulfilled this criterion with OR 2.4, 95% CI (1.0 - 6.7), p = 0.07 for mean, OR 1.7, 95% CI (1.0 - 3.2), p = 0.06 for peak, and OR 1.7, 95% CI (1.0 - 3.2), p = 0.06 for maximum change in viral load. Also, CMV viral load kinetics captured a large percentage of ganciclovir's effect by week 24—mean captured 86.3%, peak 82.7%, maximum change 94.5%, and shedding rate 93.8% (Figure 9B). Super Learner models built

using baseline characteristics of acute graft versus host disease, donor CMV serostatus, and baseline viral load plus all viral load kinetics as in the main analysis, were able to predict CMV disease by week 24 with cv-AUC greater than 75% (Figure 9C). The results of this validation procedure support not only the robustness of our findings that viral load kinetics can serve as surrogate endpoints for clinical outcomes under different treatment settings, but also the applicability to the modern antiviral prophylaxis setting.

Discussion

Our study constitutes the first analysis of viral load kinetics as surrogates for CMV clinical outcomes after HCT based on data from two randomized, placebo-controlled trials with highly successful interventions (early treatment and prophylaxis with ganciclovir). Because the data were obtained from placebo-controlled RCTs, we were able to apply the Prentice definition and consequently demonstrated that CMV viral load kinetics may be valid surrogates for CMV disease or death after HCT. Several viral load kinetics captured greater than 90% of ganciclovir's clinical effects in both treatment and prophylaxis trials. Our analysis with Super Learner shows that viral load kinetics can be used to make highly sensitive and specific predictions of clinical outcomes. In addition, in both trials, Super Learner predictions had similar accuracy when built on placebo or ganciclovir group viral load kinetics, providing additional support for the Prentice analysis. To our knowledge, our study is the first to harness the power of machine learning to evaluate virologic outcomes as surrogate endpoints. Likewise, the percentage of antiviral treatment effect captured by CMV viral load has not been estimated previously.

In this study, we used modern laboratory testing and statistical techniques to analyze frozen samples from CMV treatment and prophylaxis clinical trials performed more than 25 years ago. Because of the availability of both archived samples and clinical data from the Fred Hutch Long Term Follow Up Department, we were able to establish a direct link between viral

load suppression and improvement in clinical outcomes at extended follow up times. Because of the success of these studies, it is no longer ethical to include placebo arms in clinical trials, as patients would progress to CMV disease at much higher rates than current standards of care allow. In the treatment trial and, to a slightly lesser extent, in the prophylaxis trial, CMV disease occurred in a large percentage of patients, providing a clinical endpoint-rich environment and a dynamic range of CMV viral loads that will not be observed in any future CMV treatment trials.

We evaluated viral load kinetics, rather than viral load itself, because whereas HIV and hepatitis C viral loads respond to antiviral treatment with a stereotypic decline[28,29], CMV viral load response to treatment is more variable and depends somewhat on the immunologic status of the transplant recipient[30,31]. In fact, we demonstrated that at some time points, including absolute lymphocyte count, an indicator of CMV immune recovery[27], improved many of the predictions of clinical outcomes from the Super Learner models. Moreover, spontaneous clearance of virus in the absence of treatment is also often observed[32]. We included multiple CMV viral load kinetics in our analysis to determine which aspects of this variability correlate with clinical outcomes[32–34]. In the treatment trial, we found that mean and peak viral load are valid surrogates of both CMV disease and the composite outcome of CMV disease or death. In the prophylaxis trial, we identified percent shedding as a surrogate of CMV disease. This difference between surrogate kinetics based on treatment setting may be significant in terms of the underlying biology of CMV under treatment versus prophylaxis. In the treatment setting, the magnitude of viral load may be more predictive of tissue damage whereas the number of viral reactivations under prophylaxis may portend a higher risk of CMV disease. Considering these differences may be important in designing future antiviral trials based on viral load-based surrogate endpoints.

The main limitation of our study—that the data on which it is based have emerged from trials performed in an earlier era of HCT—is also its strength. In the treatment RCT, because patients were not treated with ganciclovir until viral cultures were positive, viral loads as

measured by CMV PCR at the time of randomization were much higher and more variable than current standards of care allow[35]. It is precisely this large range of viral loads that has allowed us to show which aspects of viral load are most predictive of CMV-related clinical outcomes. This would not be possible using data from the modern era. The existence of placebo groups in these trials demonstrated that decreasing viral load with an antiviral is the mechanism by which those in the treatment group were protected from tissue-invasive disease.

In addition, applying our methods to viral load data obtained from the prophylaxis RCT clarified that our findings are generalizable to lower viral loads. Notably, the ganciclovir prophylaxis trial design bears some remarkable similarities with the recent letermovir phase III RCT. In both trials, antivirals were given early after transplant and continued through day 100 post-HCT with clinically-significant infection and disease outcomes assessed at 24 weeks, suggesting that our findings are relevant for modern clinical practice.

However, HCT practices have changed in several important ways since the historic ganciclovir trials were conducted. Both clinical trials we analyzed included only patients who received myeloablative conditioning prior to bone marrow transplantation (BMT), and the patient population was considerably younger than modern transplant recipients[36]. In the current era, most patients receive peripheral blood stem cell (PBSCT) or umbilical cord transplants (UCT) rather than BMT. In terms of CMV infection and disease risk, PBSCT recipients more often develop CMV infection and disease in the early post-transplant period (first 100 days) than in BMT, but rates are similar later after transplant when we assessed clinical outcomes[37]. CMV infection occurs more frequently in recipients of UCT[38]. In the Goodrich et al. studies, the majority of transplants were matched-related (68% in the treatment study, 52% in the prophylaxis study); a smaller percentage were unrelated (19% in the treatment study, 23% in the prophylaxis study). In the modern era, mismatched and haploidentical transplants are increasingly common[1,2]. Recipients of mismatched transplants have higher rates of CMV disease[39] and haploidentical transplant recipients have higher peak viral loads[40]. Modern HCT

has increasingly employed alternative donors[41–43], and in these settings, recipients have higher rates of infection and disease and higher viral loads, rendering our results relevant.

In addition, all Goodrich et al. participants received myeloablative conditioning, whereas reduced-intensity conditioning regimens are now used frequently. Patients receiving reduced-intensity conditioning are less likely to have high-grade CMV infection[44], but overall rates of infection are similar[44,45]. Also, on average patients in both the treatment and prophylaxis trials were in their early 30s, and the oldest patient in either trial was 56 years old. Whereas age has not been found to be a major risk factor for CMV reactivation or disease after HCT[39,45], it is likely that age, cell source, donor match and relatedness, and conditioning regimens play a role in CMV-specific immune regulation after HCT, and we must acknowledge these limitations in our study data.

With those stated limitations and despite many changes in HCT practices, CMV disease and mortality continue to occur more frequently when viral loads rise to higher levels in all risk groups[1,2,44], supporting the validity of our study in the modern HCT setting. Using data from placebo-controlled RCTs, we show directly that reducing viral load is the mechanism through which CMV disease is reduced—a mechanism that applies to treatment and prophylaxis and both early and modern settings.

In conclusion, this study provides evidence from two placebo-controlled RCTs, using state-of-the-art statistical and machine learning techniques, that CMV viral load kinetics are valid surrogate markers for CMV disease or death in HCT recipients. These results strengthen the premise of current regulatory draft statements[46] and the recent clinical trials leading to approval of the novel agent letermovir[9,47]. CMV viral load kinetics could become powerful tools for developing and guiding the use of CMV drugs and immunotherapies for treatment or prophylaxis. In addition, our analytic approach could serve as a framework for validating surrogate markers for other viral infections, facilitating antiviral drug and immunotherapy development to eliminate viral complications after HCT.

Methods

Original study designs

Study methods for the original RCTs are included in the Supplemental Methods[17,18]. Briefly, in the early treatment study[17], CMV seropositive recipients or recipients of seropositive allogeneic bone marrow transplants for hematologic malignancies at the Fred Hutch from 1989-1990 underwent weekly CMV surveillance with viral cultures from blood, urine, and throat swabs. If any surveillance cultures were positive prior to day 80 following transplant, patients were randomized to receive ganciclovir or placebo, stratified by the presence of acute GVHD, through day 100 post-HCT. The primary endpoints were CMV disease (confirmed by biopsy or culture) and death by day 100 post-transplant. Patients were observed for outcomes until day 180. CMV disease events were defined according to established guidelines[48].

In the prophylaxis study, CMV seropositive recipients undergoing allogeneic bone marrow transplant for hematologic malignancy requiring total body irradiation or busulfan-cyclophosphamide were randomized at marrow engraftment to receive ganciclovir or placebo through day 100 after HCT or until a study endpoint of CMV infection (positive viral culture from surveillance culture), CMV disease, neutropenia, or death was reached. Additional clinical trial methods are available in the Supplemental Methods.

Extended clinical outcome analysis

We extended outcome assessment in both original RCT populations for both CMV disease and mortality to twenty years by reviewing records maintained by the Fred Hutch Long Term Follow Up Department. See additional details in the Supplemental Methods.

Viral load testing

Leftover clinical samples were stored in a repository from all patients undergoing HCT who gave their consent under a research protocol approved by the institutional review board. From this

repository, we identified frozen serum samples spaced at approximately weekly intervals from day 0 to 100 after transplantation. The University of Washington Molecular Virology Laboratory performed quantitative CMV DNA PCR testing using a laboratory-developed assay[49]. The assay's limit of quantification is 71 IU/mL; the limit of detection is 36 IU/mL. Additional information about the assay is given in the Supplemental Methods.

Viral load kinetics

We determined baseline viral loads at or near randomization and binned subsequent viral loads into weekly intervals for five weeks after randomization. Viral load data collected after diagnosis of CMV disease were removed from analysis. *VL* was defined as the log 10-converted viral load measured in IU/mL. *Maximum change in VL* was calculated by subtracting week 1 through week 5 VL from the baseline VL and finding the maximum of these values. *Mean VL* was defined as the average VL from week 1 through 5. *Peak VL* was defined as the highest VL measured from week 1 to 5. *Percentage of positive viral loads (shedding rate)* was defined as percentage of available weekly viral loads above the limit of detection. Additional details regarding the timing of viral load samples and calculation of viral load kinetics are provided in the Supplemental Methods.

Absolute lymphocyte count kinetics

We determined absolute lymphocyte count at randomization by choosing the ALC measured on the day of randomization or one day prior if randomization day ALC was not available. The peak ALC was the highest ALC from randomization to 35 days (five weeks) after randomization; ALC nadir was the lowest ALC from randomization to day 35 post-randomization; mean ALC was the average ALC from randomization to day 35 post-randomization.

Clinical endpoints

CMV disease (right-censored for death) and the first event of CMV disease or death were defined as the clinical outcomes of interest. We performed surrogate analyses on the occurrence of these endpoints by time from randomization/treatment initiation rather than time from transplantation. Thus, whereas the original study[17] defined clinical endpoints at 100 and 180 days after HCT, we defined clinical endpoints for the surrogate analysis at weeks 8, 24, and 48 after randomization (Figure 1C). For the early treatment trial, week 8 post-randomization was chosen as the first clinical outcome to approximate the RCT's study endpoint, as all of the clinical endpoints that occurred by 100 days after transplant had occurred by week 8 post-randomization. For the prophylaxis trial, patients were randomized earlier (at engraftment rather than positive viral culture), and thus, week 14 (approximately 100 days post-randomization) was chosen. For both studies 24 and 48 weeks were chosen as later endpoints to approximate 180 days and 1 year after randomization.

Statistics

All statistical analyses were performed in R (version 3.5.0)[50]. Additional information regarding the methods, including all R packages, their versions, and 'SuperLearner' libraries used, are provided in the Supplemental Methods.

Survival and cumulative incidence analysis

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. 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.

Validation of surrogate markers under the Prentice criteria

The Prentice criteria are met when a hypothesis test of the treatment effect on the surrogate endpoint is a valid hypothesis test of the treatment effect on the clinical endpoint[22]. We evaluated whether each viral load kinetic satisfied the first Prentice criterion by comparing the mean values of the viral load kinetics in the ganciclovir and placebo groups using a two-tailed student's t test. We evaluated the second Prentice criterion using logistic regression models of the association between each viral load kinetic marker and each clinical endpoint, adjusting for baseline characteristics: aGVHD, donor CMV serostatus, and randomization viral load but not treatment group. We evaluated the third Prentice criterion using logistic regression models of the association between each VL kinetic marker and each clinical endpoint, adjusting for treatment group and baseline characteristics: aGVHD, donor CMV serostatus, and randomization viral load. The second Prentice criterion was satisfied if the coefficient of the viral load kinetic term was significantly different from zero, indicating a significant association between the clinical endpoint and viral load kinetic. The third Prentice criterion was satisfied if the coefficient of the treatment assignment term was close to zero ($p \geq 0.20$), i.e., when holding the value of the viral load kinetic constant, the outcome was not more likely to occur in one of the treatment groups. In addition, if there was evidence of effect modification between a viral load kinetic and treatment group ($p < 0.20$ in a logistic regression model containing an interaction term), the third criterion was not satisfied.

Percentage of treatment effect captured

We quantified how much of ganciclovir's effect on clinical outcomes could be attributed to its effect on viral load kinetics. This quantity is called the proportion of treatment effect captured by

the surrogate (PCS)[23]. A PCS > 63% is considered "moderate;" PCS > 85% is "substantial"; PCS > 93% is "almost perfect"[23]. Details regarding the PCS method and our implementation are provided in the Supplemental Methods.

Super Learner ensemble machine learning

Super Learner is an ensemble machine learning method that estimates a conditional outcome risk model as the optimal combination of individual regression algorithms that maximize a cross-validated criterion for best disease classification accuracy. Specifically, we minimized the leave-one-out, cross-validated area under the receiver operating curve (cv-AUC). Super Learner prediction models were built with the same baseline covariates and viral load kinetics defined for the logistic regression analysis and were fit on data from the placebo group alone, the ganciclovir group alone, and the combined treatment groups, with individual regression algorithms specified in the Supplemental Methods. cv-AUCs were calculated for each Super Learner prediction model, with a pre-defined benchmark that cv-AUCs greater than 85% would provide evidence for the fitted values (i.e., predicted outcome risks) as potentially valid surrogates. Super Learning was implemented using the R package 'SuperLearner'[51].

Data sharing

De-identified individual participant data may be requested for further research from the corresponding author.

Study approval

The original studies were approved by the Fred Hutch institutional review board and the Food and Drug Administration. All patients or their legal guardians provided informed consent. The viral load surrogate study was also approved by the Fred Hutch institutional review board.

Contributors

MJB, JTS, PBG conceived of the study design. TSA, CW, MH, MEF, KRJ, LC, MJB contributed to data collection. ERD, BDW, BB, PBG, NC, HW developed the data analysis plan. ERD, BDW, BB, JLG performed data analysis and modeling. ERD, NC, HW, TCM, MM, PBG, JTS, MJB interpreted the data and its analysis. ERD developed the figures. All authors participated in drafting and review of the manuscript.

Acknowledgments

We dedicate this work to the late Drs Joel D Meyers and James M Goodrich. We thank the original study participants for their commitment to research and research staff for their careful work. Also, we would like to thank Jo Tono, Vera Okolo, Heather Andrews, Darneshia Smith, and Laurel Joncas Schronce for sample management; the late John Hansen for maintaining the Research Cell Bank serum collection, which provided additional samples to the IDS Repository; Elizabeth Nguyen, Lisa Chung, Sonia Goyal, and Louise Kimball for record review; Bryan Mayer for biostatistical and R programming advice; Florencia Tettamanti-Boshier, Emily Ford, Hannah Imlay, Daniel Reeves, David Swan, and E. Fabian Cardozo-Ojeda for figure development advice; Hillary Cuesta and Alexandria Kachurak for research and manuscript support.

Funding for this project was provided by a collaborative research grant from Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA, the US National Institutes of Health: KL2 TR002317 (E.D.), K24 HL093294 (M.B.), CA 15704, and the Fred Hutch Vaccine and Infectious Disease Division (Biorepository).

References

1. Teira P, Battiwalla M, Ramanathan M, et al. Early cytomegalovirus reactivation remains associated with increased transplant-related mortality in the current era: A CIBMTR analysis. *Blood*. 2016;127(20):2427-2438. doi:10.1182/blood-2015-11-679639
2. Green ML, Leisenring W, Xie H, et al. Cytomegalovirus viral load and mortality after haemopoietic stem cell transplantation in the era of pre-emptive therapy: a retrospective cohort study. *Lancet Haematol*. 2016;3(3):e119-e127. doi:10.1016/S2352-3026(15)00289-6
3. Cantoni N, Hirsch HH, Khanna N, et al. Evidence for a bidirectional relationship between cytomegalovirus replication and acute graft-versus-host disease. *Biol Blood Marrow Transplant*. 2010;16(9):1309-1314. doi:10.1016/j.bbmt.2010.03.020
4. Nichols WG, Corey L, Gooley T, Davis C, Boeckh M. High Risk of Death Due to Bacterial and Fungal Infection among Cytomegalovirus (CMV)– Seronegative Recipients of Stem Cell Transplants from Seropositive Donors: Evidence for Indirect Effects of Primary CMV Infection. *J Infect Dis*. 2002;185:273-282.
5. Young J-AH, Logan BR, Wu J, et al. Infections after Transplantation of Bone Marrow or Peripheral Blood Stem Cells from Unrelated Donors. *Biol Blood Marrow Transplant*. 2016;22(2):359-370. doi:10.1016/j.bbmt.2015.09.013
6. Marty FM, Ljungman P, Papanicolaou GA, et al. Maribavir prophylaxis for prevention of cytomegalovirus disease in recipients of allogeneic stem-cell transplants: A phase 3, double-blind, placebo-controlled, randomised trial. *Lancet Infect Dis*. 2011;11(4):284-292. doi:10.1016/S1473-3099(11)70024-X
7. Snydman DR. Why did maribavir fail in stem-cell transplants? *Lancet Infect Dis*. 2011;11(4):255-257. doi:10.1016/S1473-3099(11)70033-0
8. Natori Y, Alghamdi A, Tazari M, et al. Use of Viral Load as a Surrogate Marker in Clinical Studies of Cytomegalovirus in Solid Organ Transplantation: A Systematic Review and

Meta-analysis. *Clin Infect Dis*. 2018;66(4):617-631. doi:10.1093/cid/cix793

9. Marty FM, Ljungman P, Chemaly RF, et al. Letermovir Prophylaxis for Cytomegalovirus in Hematopoietic-Cell Transplantation. *N Engl J Med*. 2017;377(25):2433-2444. doi:10.1056/NEJMoa1706640

10. Petrella M, Montaner J, Batist G, Wainberg MA. The Role of Surrogate Markers in the Clinical Development of Antiretroviral Therapy: A Model for Early Evaluation of Targeted Cancer Drugs. *Cancer Invest*. 2004;22(1):149-160. doi:10.1081/CNV-120027590

11. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996;125(7):605-613. doi:10.7326/0003-4819-125-7-199610010-00011

12. De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials: Summary of a National Institutes of Health Workshop. In: *Controlled Clinical Trials*. Vol 22. ; 2001:485-502. doi:10.1016/S0197-2456(01)00153-2

13. FDA, Cder, Mccrayk. Chronic Hepatitis C Virus Infection: Developing Direct-Acting Antiviral Drugs for Treatment Guidance for Industry. *www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidance*. 2017;(November). doi:https://dx.doi.org/10.1353/hpu.2014.0156

14. Chen J, Florian J, Carter W, et al. Earlier sustained virologic response end points for regulatory approval and dose selection of hepatitis C therapies. *Gastroenterology*. 2013;144(7):1450-1455.e2. doi:10.1053/j.gastro.2013.02.039

15. Murray JS, Elashoff MR, Iacono-Connors LC, Cvetkovich TA, Struble KA. The use of plasma HIV RNA as a study endpoint in efficacy trials of antiretroviral drugs. *AIDS*. 1999;13(7):797-804. doi:10.1097/00002030-199905070-00008

16. O'Brien WA, Hartigan PM, Daar ES, Simberkoff MS, Hamilton JD. Changes in plasma HIV RNA levels and CD4+ lymphocyte counts predict both response to antiretroviral therapy and therapeutic failure. *Ann Intern Med*. 1997;126(12):939-945. doi:10.7326/0003-4819-126-12-199706150-00002

17. 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(23):1601-1607. doi:10.1056/NEJM199112053252303
18. 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(3):173-178. doi:10.7326/0003-4819-118-3-199302010-00003
19. Schmidt GM, Horak DA, Niland JC, Duncan SR, Forman SJ, Zaia JA. A randomized, controlled trial of prophylactic ganciclovir for cytomegalovirus pulmonary infection in recipients of allogeneic bone marrow transplants; The City of Hope-Stanford-Syntex CMV Study Group. *N Engl J Med*. 1991;324(15):1005-1011. doi:10.1056/NEJM199104113241501
20. Dieterich DT, Kotler DP, Busch DF, et al. Ganciclovir treatment of cytomegalovirus colitis in AIDS: a randomized, double-blind, placebo-controlled multicenter study. *J Infect Dis*. 1993;167(2):278-282.
21. Winston DJ, Ho WG, Bartoni K, et al. Ganciclovir prophylaxis of cytomegalovirus infection and disease in allogeneic bone marrow transplant recipients. Results of a placebo-controlled, double-blind trial. *Ann Intern Med*. 1993;118(3):179-184.
22. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*. 1989;8(4):431-440. doi:10.1016/0197-2456(85)90012-1
23. Kobayashi F, Kuroki M. A new proportion measure of the treatment effect captured by candidate surrogate endpoints. *Stat Med*. 2014;33(19):3338-3353. doi:10.1002/sim.6180
24. Van Der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6(1). doi:10.2202/1544-6115.1309
25. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): A population-based study. *Lancet Respir Med*. 2015;3(1):42-52. doi:10.1016/S2213-

2600(14)70239-5

26. Gilbert PB, Luedtke AR. Statistical Learning Methods to Determine Immune Correlates of Herpes Zoster in Vaccine Efficacy Trials. *J Infect Dis*. 2018;218(2):S99-S101. doi:10.1093/infdis/jiy421

27. Gardiner BJ, Nierenberg NE, Chow JK, Ruthazer R, Kent DM, Snydman DR. Absolute lymphocyte count: A predictor of recurrent cytomegalovirus disease in solid organ transplant recipients. *Clin Infect Dis*. 2018;67(9):1395-1402. doi:10.1093/cid/ciy295

28. Perelson AS, Essunger P, Cao Y, et al. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*. 1997;387(6629):188-191. doi:10.1038/387188a0

29. Neumann AU, Lam NP, Dahari H, et al. Hepatitis C Viral Dynamics In vivo and the Antiviral Efficacy of Interferon-α Therapy. *Science (80-)*. 1998;282(5386):103-107.

30. Nichols WG, Corey L, Gooley T, et al. Rising pp65 antigenemia during preemptive anticytomegalovirus therapy after allogeneic hematopoietic stem cell transplantation: Risk factors, correlation with DNA load, and outcomes. *Blood*. 2001;97(4):867-874. doi:10.1182/blood.V97.4.867

31. Tey S-K, Kennedy GA, Cromer D, et al. Clinical Assessment of Anti-Viral CD8+ T Cell Immune Monitoring Using QuantiFERON-CMV® Assay to Identify High Risk Allogeneic Hematopoietic Stem Cell Transplant Patients with CMV Infection Complications. *PLoS One*. 2013;8(10):e74744. doi:10.1371/journal.pone.0074744

32. Camargo JF, Kimble E, Rosa R, et al. Impact of Cytomegalovirus Viral Load on Probability of Spontaneous Clearance and Response to Preemptive Therapy in Allogeneic Stem Cell Transplantation Recipients. *Biol Blood Marrow Transplant*. 2018;24(4):806-814. doi:10.1016/j.bbmt.2017.11.038

33. Muñoz-Cobo B, Solano C, Costa E, et al. Dynamics of cytomegalovirus (CMV) plasma DNAemia in initial and recurrent episodes of active CMV infection in the allogeneic stem

cell transplantation setting: Implications for designing preemptive antiviral therapy strategies. *Biol Blood Marrow Transplant*. 2011;17(11):1602-1611. doi:10.1016/j.bbmt.2011.08.014

34. Buyck HCE, Griffiths PD, Emery VC. Human cytomegalovirus (HCMV) replication kinetics in stem cell transplant recipients following anti-HCMV therapy. *J Clin Virol*. 2010;49(1):32-36. doi:10.1016/j.jcv.2010.06.018

35. Einsele H, Ehninger G, Hebart H, et al. Polymerase chain reaction monitoring reduces the incidence of cytomegalovirus disease and the duration and side effects of antiviral therapy after bone marrow transplantation. *Blood*. 1995;86(7):2815-2820. doi:10.1182/blood.v86.7.2815.bloodjournal8672815

36. McDonald GB, Sandmaier BM, Mielcarek M, et al. Survival, nonrelapse mortality, and relapse-related mortality after allogeneic hematopoietic cell transplantation: Comparing 2003–2007 versus 2013–2017 cohorts. *Ann Intern Med*. 2020;172(4):229-239. doi:10.7326/M19-2936

37. Guerrero A, Riddell SR, Storek J, et al. Cytomegalovirus Viral Load and Virus-Specific Immune Reconstitution after Peripheral Blood Stem Cell versus Bone Marrow Transplantation. *Biol Blood Marrow Transplant*. 2012;18(1):66-75. doi:10.1016/j.bbmt.2011.05.010

38. Melendez-Munoz R, Marchalik R, Jerussi T, et al. Cytomegalovirus Infection Incidence and Risk Factors Across Diverse Hematopoietic Cell Transplantation Platforms Using a Standardized Monitoring and Treatment Approach: A Comprehensive Evaluation from a Single Institution. *Biol Blood Marrow Transplant*. 2019;25(3):577-586. doi:10.1016/j.bbmt.2018.10.011

39. Green ML, Leisenring W, Stachel D, et al. Efficacy of a Viral Load-Based, Risk-Adapted, Preemptive Treatment Strategy for Prevention of Cytomegalovirus Disease after Hematopoietic Cell Transplantation. *Biol Blood Marrow Transplant*. 2012;18(11):1687-

1699. doi:10.1016/j.bbmt.2012.05.015

40. Huntley D, Gimenez E, Pascual M, Hernandez-Boluda J, Gago B, Vazquez L. Incidence, features and outcomes of Cytomegalovirus DNAemia in unmanipulated haploidentical allogeneic hematopoietic stem cell transplantation with post-transplantation cyclophosphamide. *Transpl Infect Dis*. 2019. doi:10.1111/tid.13206

41. Luznik L, O'Donnell P V., Symons HJ, et al. HLA-Haploidentical Bone Marrow Transplantation for Hematologic Malignancies Using Nonmyeloablative Conditioning and High-Dose, Posttransplantation Cyclophosphamide. *Biol Blood Marrow Transplant*. 2008;14(6):641-650. doi:10.1016/j.bbmt.2008.03.005

42. Gutman JA, Ross K, Smith C, et al. Chronic graft versus host disease burden and late transplant complications are lower following adult double cord blood versus matched unrelated donor peripheral blood transplantation. *Bone Marrow Transplant*. 2016;51(12):1588-1593. doi:10.1038/bmt.2016.186

43. Bashey A, Solomon SR. T-cell replete haploidentical donor transplantation using post-transplant CY: An emerging standard-of-care option for patients who lack an HLA-identical sibling donor. *Bone Marrow Transplant*. 2014;49(8):999-1008. doi:10.1038/bmt.2014.62

44. Nakamae H, Kirby KA, Sandmaier BM, et al. Effect of Conditioning Regimen Intensity on CMV Infection in Allogeneic Hematopoietic Cell Transplantation. *Biol Blood Marrow Transplant*. 2009;15(6):694-703. doi:10.1016/j.bbmt.2009.02.009

45. Ljungman P, Hakki M, Boeckh M. Cytomegalovirus in Hematopoietic Stem Cell Transplant Recipients. *Hematol Oncol Clin North Am*. 2011;25(1):151-169. doi:10.1016/j.hoc.2010.11.011

46. *Cytomegalovirus in Transplantation: Developing Drugs to Treat or Prevent Disease - Guidance for Industry*.; 2018. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guida

nces/UCM608059.pdf.

47. Chemaly RF, Ullmann AJ, Stoelben S, et al. Letermovir for cytomegalovirus prophylaxis in hematopoietic-cell transplantation. *N Engl J Med*. 2014;370(19):1781-1789. doi:10.1056/NEJMoa1309533

48. Ljungman P, Boeckh M, Hirsch HH, et al. Definitions of cytomegalovirus infection and disease in transplant patients for use in clinical trials. *Clin Infect Dis*. 2017;64(1):87-91. doi:10.1093/cid/ciw668

49. Boeckh M, Huang M, Ferrenberg J, et al. Optimization of Quantitative Detection of Cytomegalovirus DNA in Plasma by Real-Time PCR. *J Clin Microbiol*. 2004;43(3):1142-1148. doi:10.1128/JCM.42.3.1142

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

51. 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.

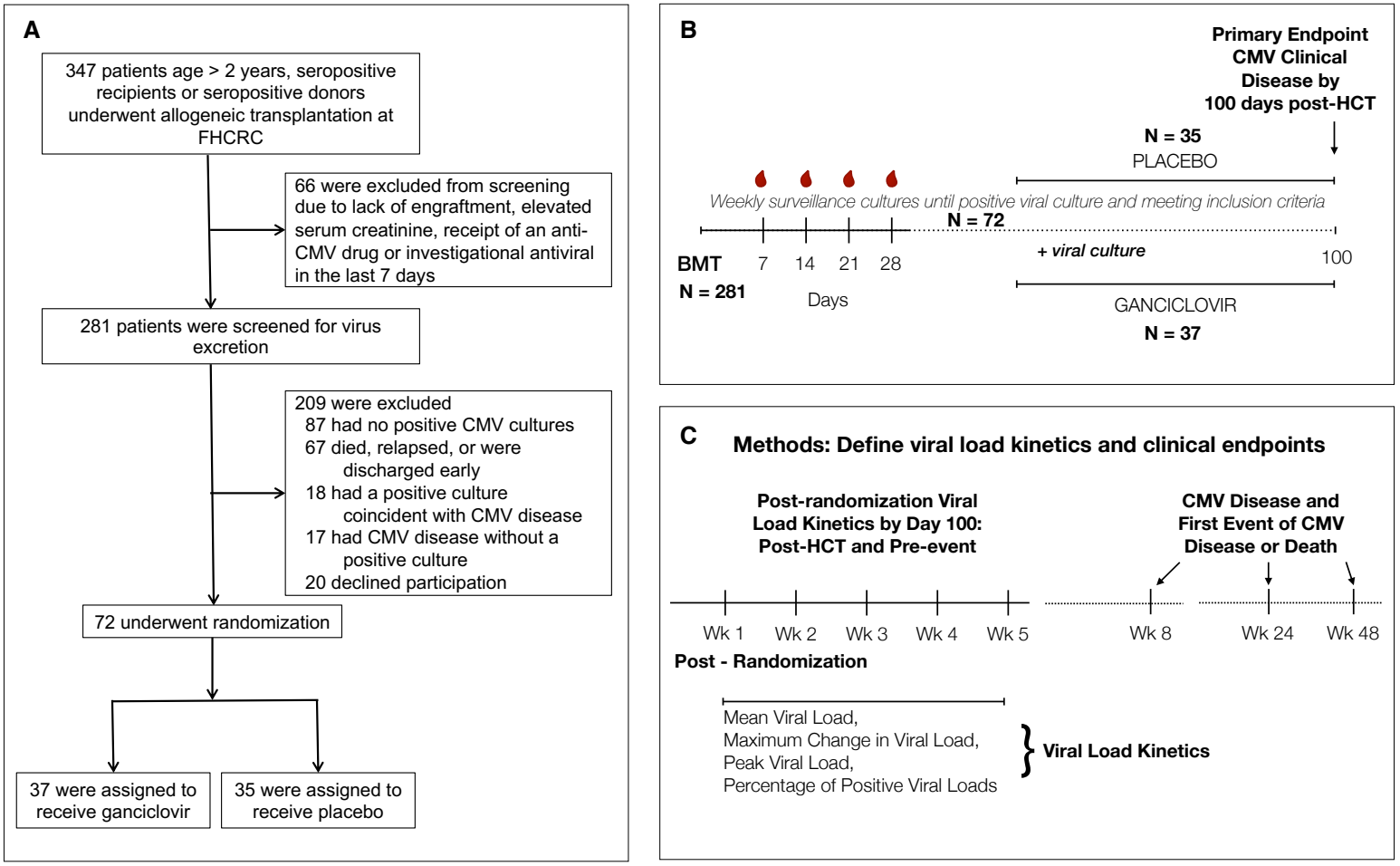

Figure 1 – **CONSORT diagram and study designs for the early treatment trial.** Study design for Goodrich et al. NEJM 1991 RCT (A & B) and for viral load kinetic analysis (C). (A) is the reconstructed CONSORT diagram for the original RCT. (B) illustrates the original study design with surveillance and screening beginning at HCT and randomization beginning at the time of first positive surveillance culture. (C) depicts the viral load kinetics study design with analysis beginning at randomization (receipt of study drug) and ending at day 100 post-HCT or a study endpoint of CMV disease or death, whichever occurred first.

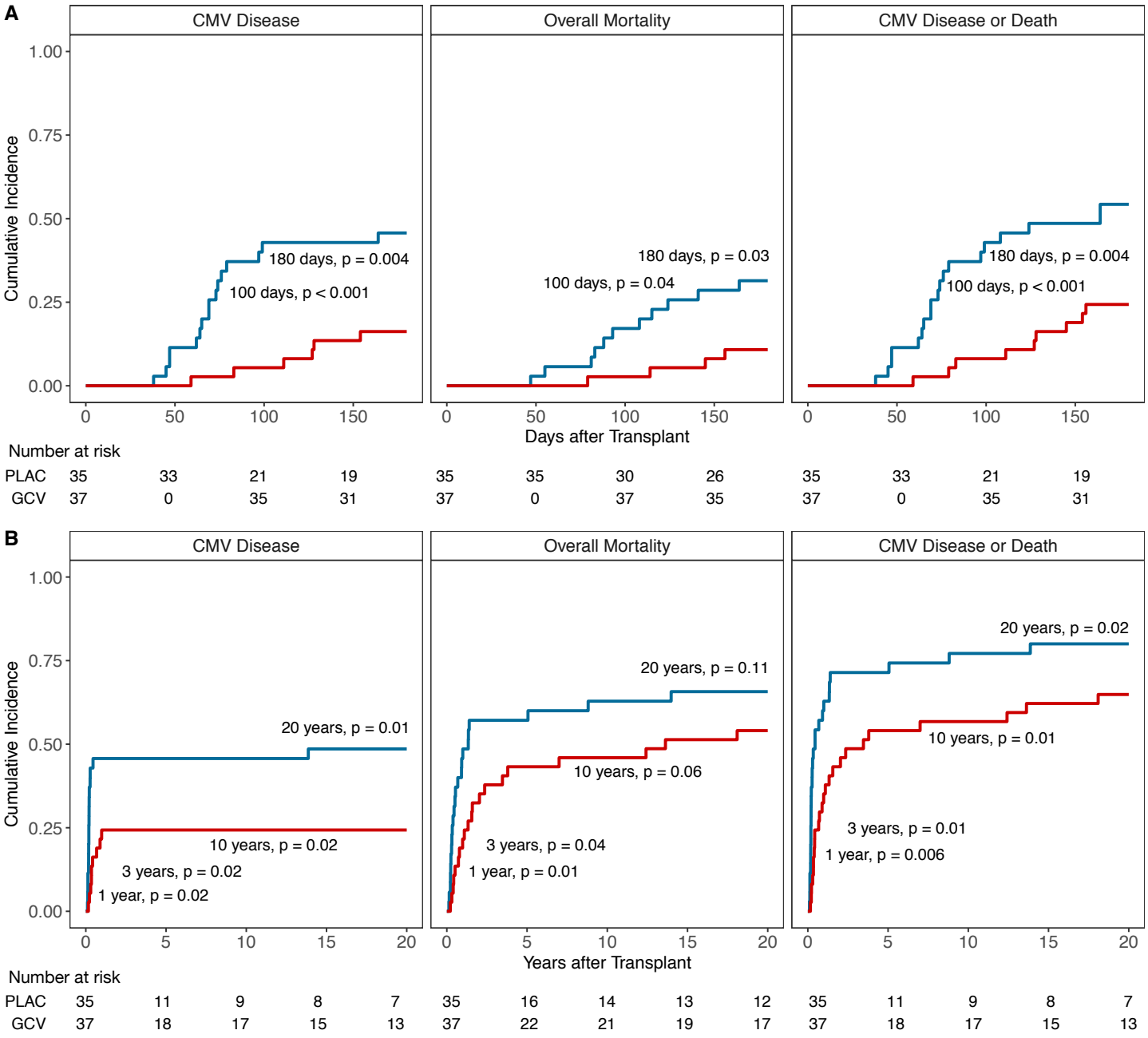

Figure 2 – **CMV disease and death clinical outcomes in the early treatment trial.** CMV disease (right-censored for death), overall mortality, and first event of CMV disease or mortality in the placebo and ganciclovir groups at time points defined in the original study (A) and at extended follow up times out to 20 years (B). In all plots, the ganciclovir group is shown in red; the placebo group is shown in blue. Numbers at risk are shown below their respective plots (PLAC indicates the placebo group. GCV indicates the treatment group). 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.

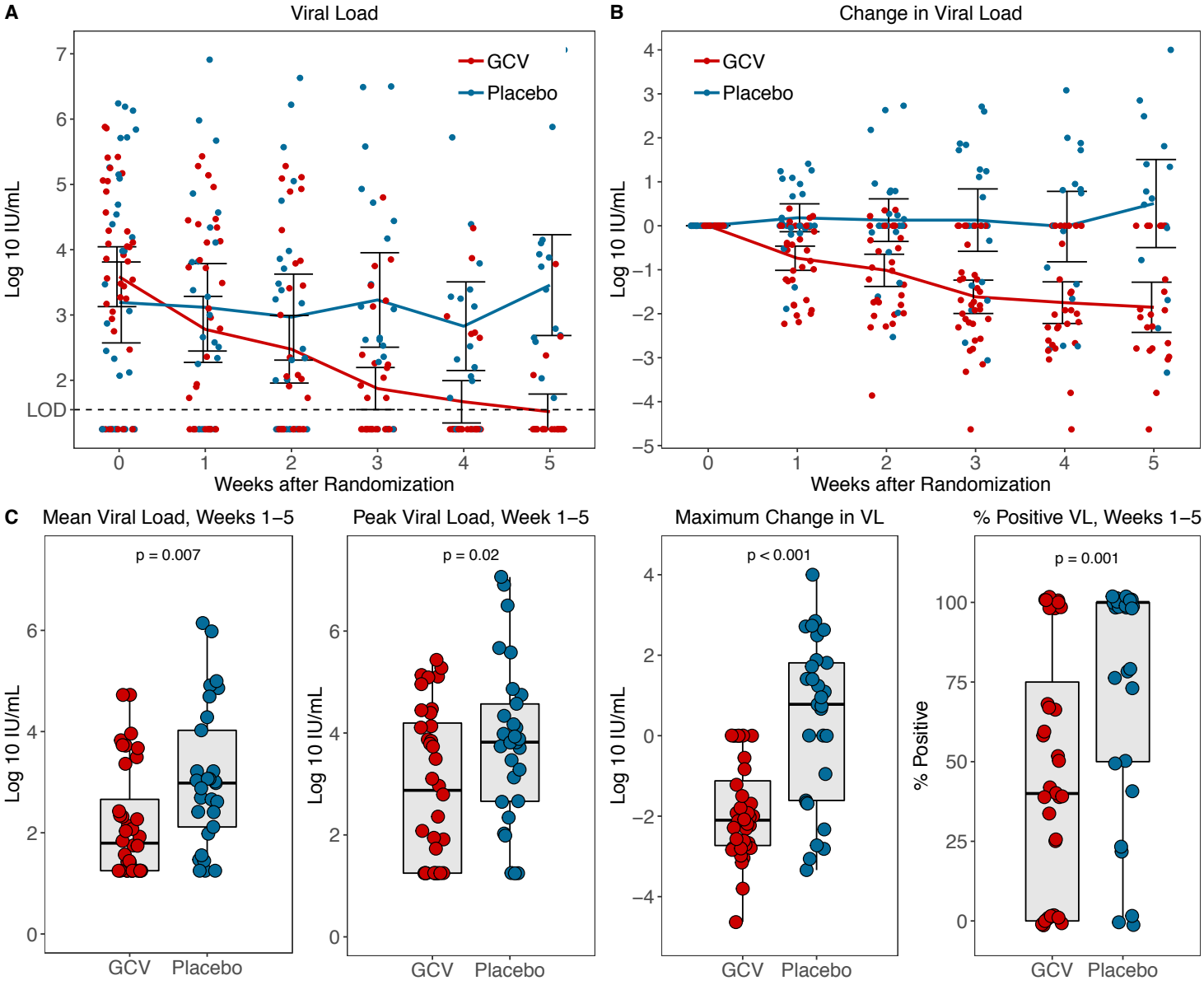

Figure 3 – **Weekly CMV viral load (VL) kinetics in the early treatment trial.** CMV viral load kinetics from time of randomization (Week 0). In A & B, VL data are shown for patients who had not reached an endpoint of CMV disease or death by that week. GCV indicates patients in the ganciclovir treatment group who are shown in red. Placebo indicates patients in the placebo treatment group who are shown in blue. Error bars indicate 95% confidence intervals. The dashed horizontal line represents the limit of detection (LOD) of the CMV viral load assay. VL kinetics summary calculations (C) were performed with the data shown in A & B. Box and whisker plots show the middle 50% of VL kinetics in grey boxes with a horizontal black line at the median. Whiskers extend upward from the third quartile at the top of the box to 1.5 times the interquartile range (the distance between first and third quartiles) and downward from the first quartile at the bottom of the box to 1.5 times the interquartile range. p-values were calculated from two-tailed t tests comparing the means of the viral kinetics in GCV versus placebo groups.

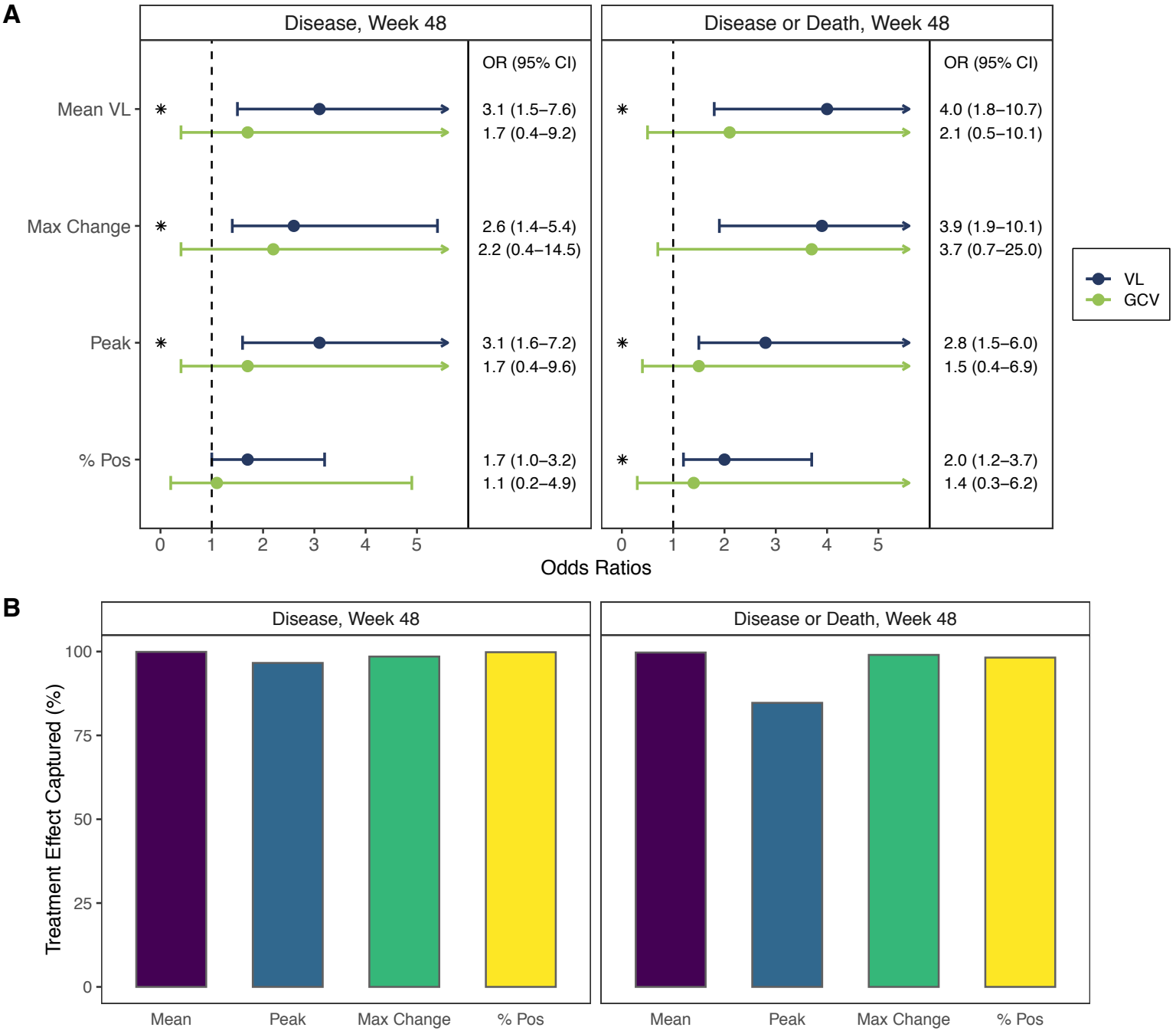

Figure 4 - **Prentice criteria (PC) evaluation using multivariate logistic regression and proportion of treatment effect captured in the early treatment trial.** (A) Forest plots of the odds ratios (OR) for associations of VL kinetics with risk for CMV disease and CMV disease or death by week 48 after randomization were calculated from logistic regression models adjusted for baseline characteristics and treatment group. OR for VL kinetics are indicated by navy dots surrounded by 95% confidence intervals (CI) indicated with navy lines; OR with 95% CI for treatment group assignment shown with light-green dots and lines. Asterisks (*) indicate viral load kinetics that met the PC by multivariable logistic regression testing, i.e. the coefficient for VL kinetic was significantly different from zero ($p < 0.05$), whereas the treatment group assignment coefficient was not significantly different from zero ($p \geq 0.20$). The treatment by marker interaction coefficient was not significantly different from zero ($p \geq 0.20$) for any kinetic. % Pos did not meet PC for CMV disease with $p = 0.07$ for VL kinetic association. Max change did not meet PC for CMV disease with $p = 0.14$ for GCV association. For Mean, Max Change, and Peak, ORs were calculated as the ratio of odds of the clinical outcome in groups differing by log 10 IU/mL. For % Pos, the OR was calculated as the ratio of odds of the clinical outcome in groups differing by 25% in percentage of samples with detectable VL. Dashed vertical lines indicate OR = 1. (B) The percentages of ganciclovir's effect on clinical outcomes captured by the candidate surrogate were calculated using Kobayashi and Kuroki's measure[12] and are shown for each of the viral load kinetics.

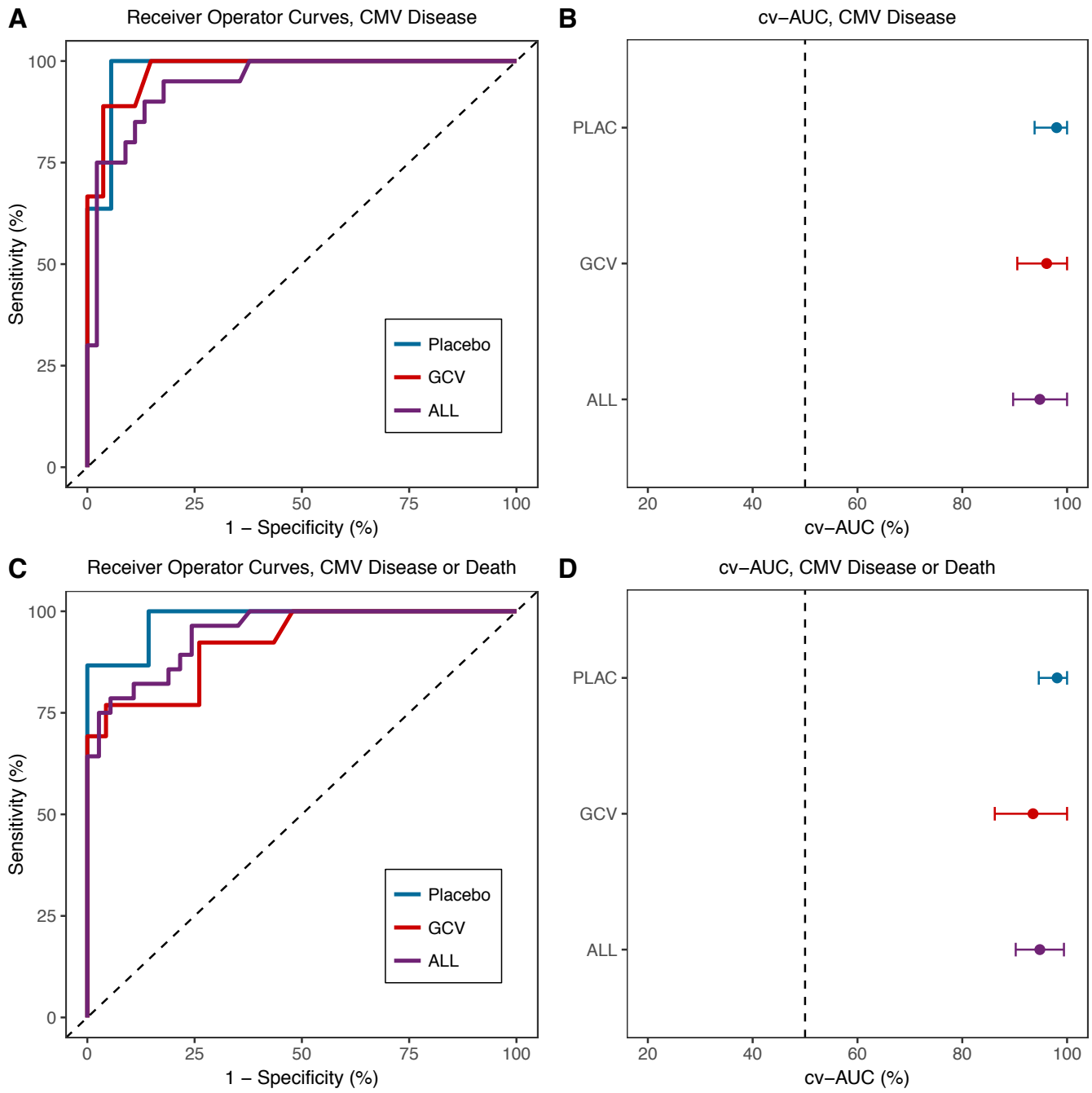

Figure 5 - **Prediction accuracy for clinical outcomes with Super Learners in the early treatment trial.** (A & C) Receiver operating characteristic curves (ROC) are shown for Super Learner predictions for CMV disease and CMV disease or death by 48 weeks after randomization. The diagonal line drawn at y = x indicates the boundary above which ROC curves describe a prediction that is better than chance. (B & D) 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. For A-D, predictions made only on data from the placebo group are in blue, from the ganciclovir group (GCV) in red, and from both treatment groups (ALL) in purple. In B & D, the vertical line indicates cv-AUC = 50%, the area under the diagonal line in A &C.

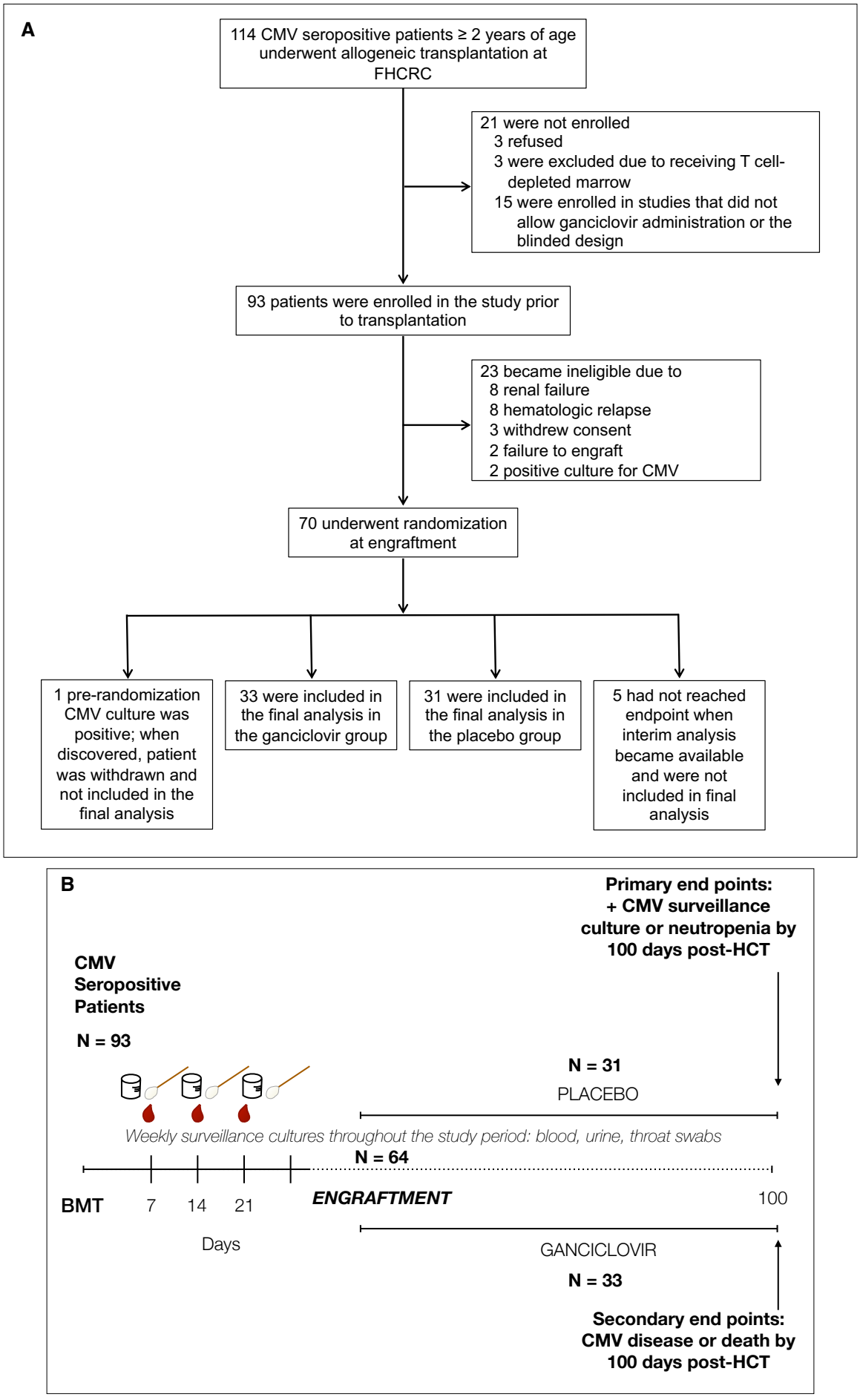

Figure 6 – CONSORT diagram and study design for the prophylaxis trial. Study design for Goodrich et al. AIM 1993 RCT. (A) is the reconstructed CONSORT diagram for the original RCT. (B) illustrates the original study design with surveillance and screening beginning at HCT and randomization beginning at the time of engraftment.

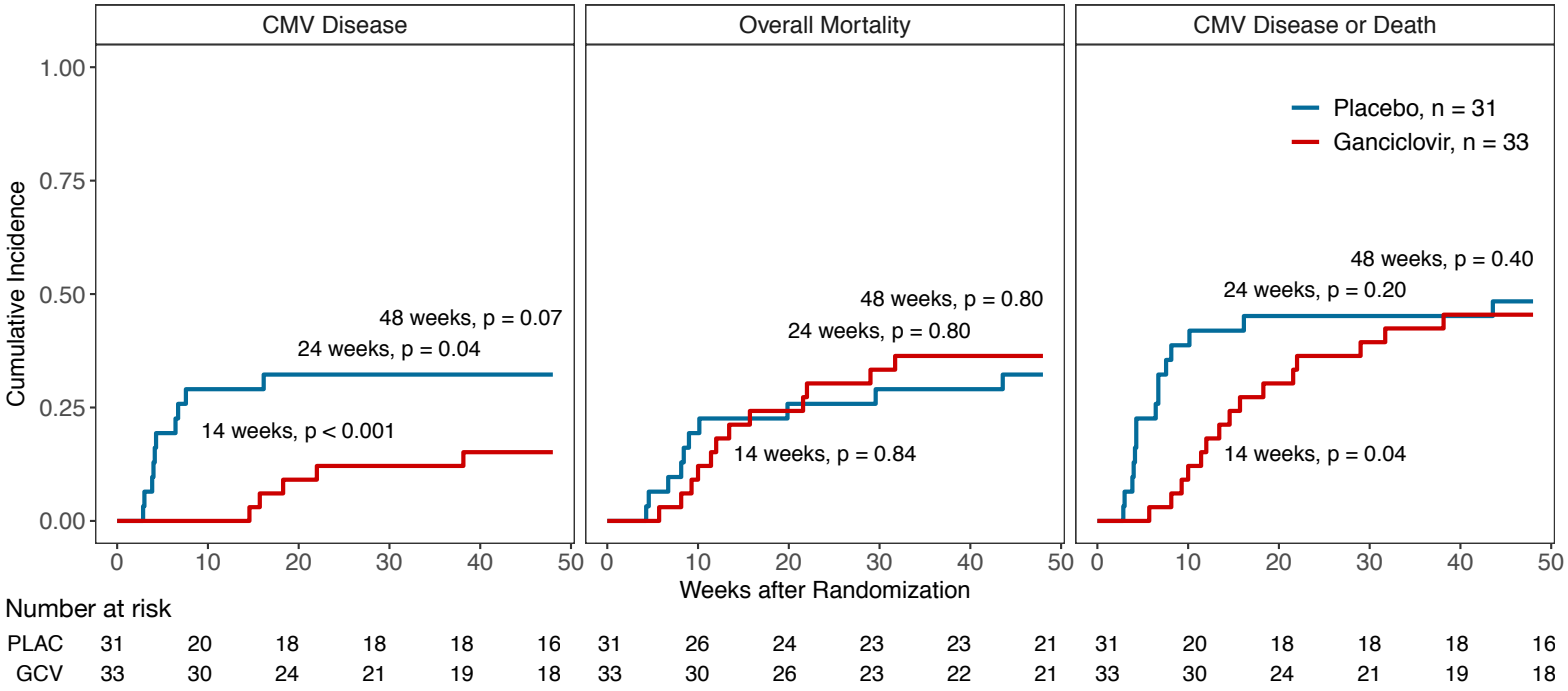

Figure 7 – **CMV disease clinical outcomes in the prophylaxis trial.** CMV disease (right-censored for death), overall mortality, and first event of CMV disease or mortality in the placebo and ganciclovir groups at 14, 24, and 48 weeks after randomization (A). The ganciclovir group is shown in red; the placebo group is shown in blue. Numbers at risk are shown below their respective plots (PLAC indicates the placebo group. GCV indicates the treatment group). 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.

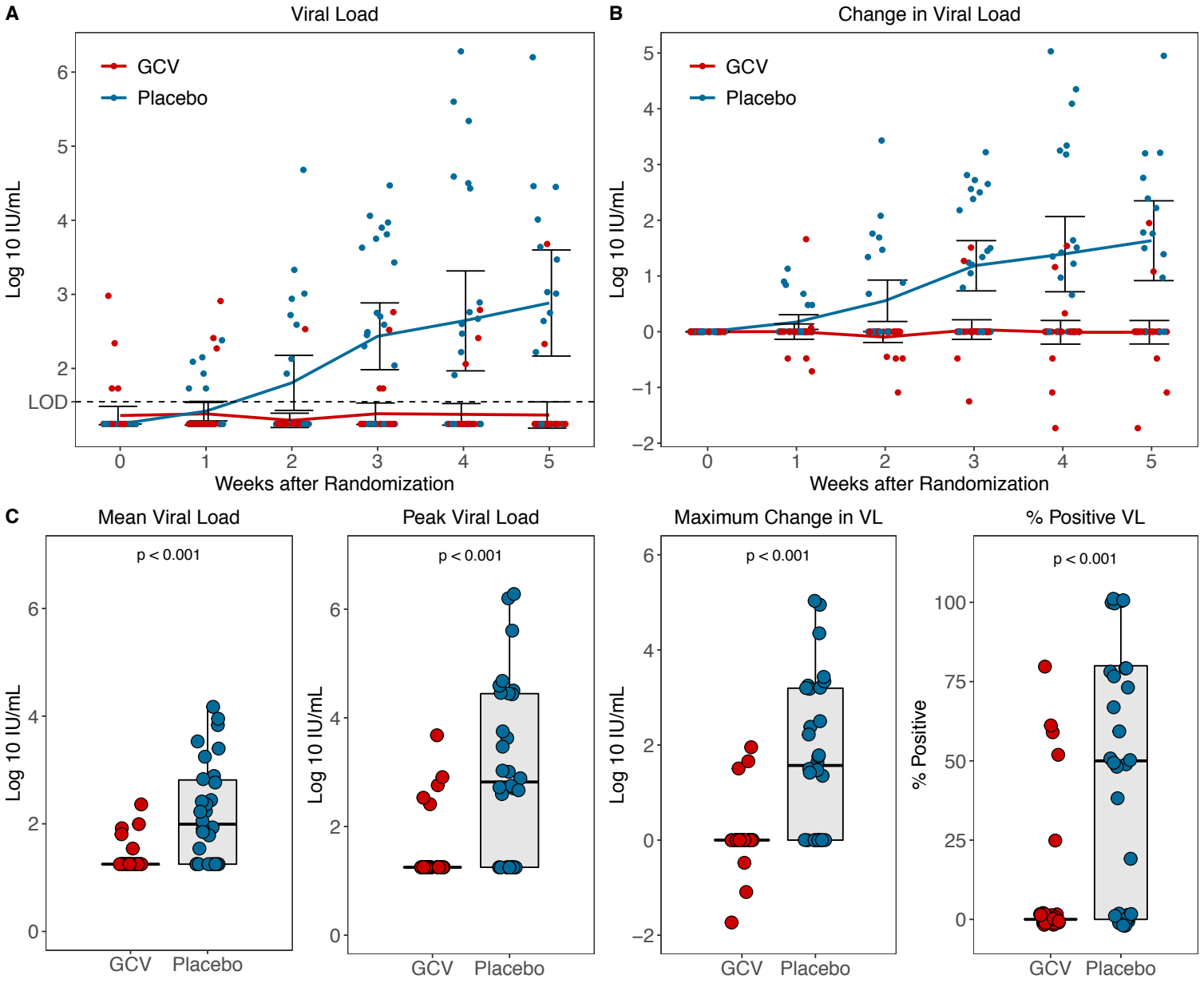

Figure 8 – **Weekly CMV viral load (VL) kinetics in the prophylaxis trial.** CMV viral load kinetics from time of randomization (Week 0). In A & B, VL data are shown for patients who had not reached an endpoint of CMV disease or death by that week. GCV indicates patients in the ganciclovir treatment group who are shown in red. Placebo indicates patients in the placebo treatment group who are shown in blue. Error bars indicate 95% confidence intervals. The dashed horizontal line represents the limit of detection (LOD) of the CMV viral load assay. VL kinetics summary calculations (C) were performed with the data shown in A & B. Box and whisker plots show the middle 50% of VL kinetics in grey boxes with a horizontal black line at the median. Whiskers indicate 1.5 times the interquartile range of the VL kinetics. p-values were calculated from two-tailed t tests comparing the means of the viral kinetics in ganciclovir (GCV) versus placebo groups.

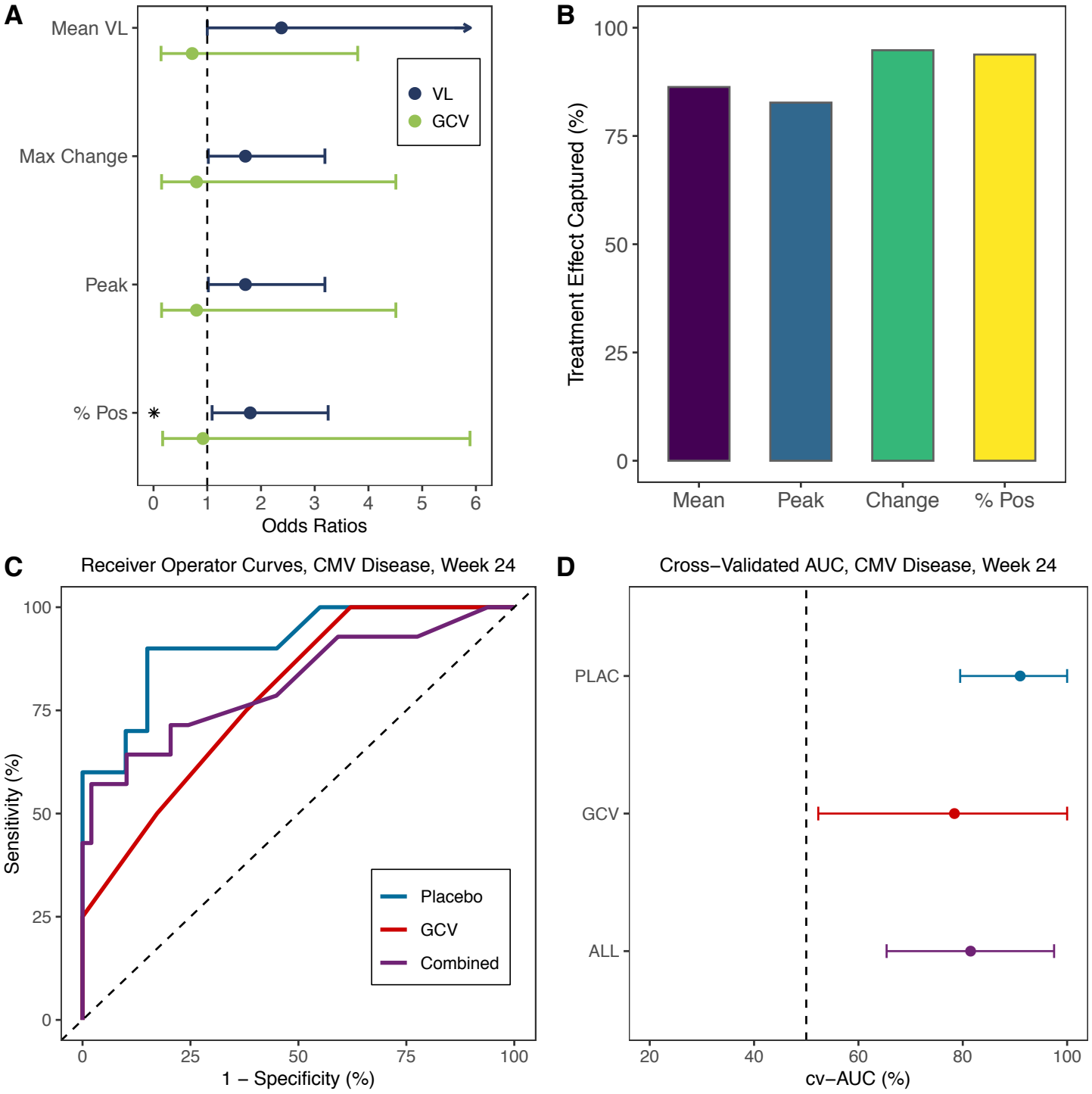

Figure 9 - **Prentice criteria (PC) evaluation using multivariate logistic regression, proportion of treatment effect captured, prediction accuracy for clinical outcomes with Super Learners in the prophylaxis trial.** (A) Forest plots of the odds ratios (OR) for associations of VL kinetics with risk for CMV disease and CMV disease or death by week 24 after randomization were calculated from logistic regression models adjusted for baseline characteristics and treatment group. OR for VL kinetics are indicated by navy dots surrounded by 95% confidence intervals (CI) indicated with navy lines; OR with 95% CI for treatment group assignment shown with light-green dots and lines. Asterisks (*) indicate viral load kinetics that met the PC by multivariable logistic regression testing. The dashed vertical line indicates OR = 1. (B) The percentages of ganciclovir's effect on clinical outcomes captured by the candidate surrogate were calculated using Kobayashi and Kuroki's measure[12] and are shown for each of the viral load kinetics indicated. (C) Receiver operating characteristic curves (ROC) are shown for Super Learner predictions for CMV disease by week 24 after randomization. The diagonal line drawn at y = x indicates the boundary above which ROC curves describe a prediction that is better than chance. (D) The Forest plot shows cross-validated area under the receiver operator curves (cv-AUC) of Super Learner predictions for CMV disease. The vertical line indicates cv-AUC = 50%, the area under the diagonal line in C. For C & D, predictions made only on data from the placebo group are in blue, from the ganciclovir group (GCV) in red, and from both treatment groups (ALL) in purple.