Supplementary Appendix A: Waveform Measure Medians

Waveform measure median values are reported by shock outcome and CPR state (Table S1) and by device type (Table S2).

Table S1: Median AMSA and MS waveform measure values from the study group are presented stratified by CPR state and shock outcome (N=692 patients, n=1372 CPR segments from shocks 2-4, n=1283 CPR-free segments from shocks 2-4). (AMSA = Amplitude Spectrum Area; CPR = Cardiopulmonary Resuscitation; IQR = Interquartile Range; MS = Median Slope; ROR = Return of Organized Rhythm)

	With CPR		Without CPR	
Shock Outcome	**ROR Median [IQR], n=801**	**No ROR Median [IQR], n=571**	**ROR Median [IQR], n=750**	**No ROR Median [IQR], n=533**
AMSA	7.15 [4.95, 9.35] (mV-Hz)	5.21 [3.26, 7.16] (mV-Hz)*	6.00 [4.02, 7.97] (mV-Hz)	3.74 [2.21, 5.27] (mV-Hz)*
MS	4.11 [2.74, 5.48] (mV/s)	2.82 [1.74, 3.90] (mV/s)*	3.47 [2.22, 4.73] (mV/s)	2.02 [1.10, 2.94] (mV/s)*

**Median significantly less than ROR (p<0.001, two-sided Wilcoxon rank-sum test)*

Table S2: Median AMSA and MS waveform measure values are presented by defibrillator model for all segments in study group (N=692 patients, n=1372 CPR segments from shocks 2-4, n=1283 CPR-free segments from shocks 2-4). Device filter bandwidths are reported based on values specified in device user manuals for EMS Mode (MRx) or Paddles ECG (LP12/LP15). (AMSA = Amplitude Spectrum Area; AUC = Area under Receiver Operating Characteristic Curve; CI = Confidence Interval; CPR = Cardiopulmonary Resuscitation; IQR = Interquartile Range; MS = Median Slope)

Device (Bandwidth)	With CPR		Without CPR	
	n	**Median [IQR]**	**n**	**Median [IQR]**
MRx (1-30 Hz)*	578		587	
AMSA		5.69 [3.54, 7.84] (mV-Hz)		4.80 [2.84, 6.77] (mV-Hz)
MS		3.23 [1.93, 4.52] (mV/s)		2.71 [1.48, 3.94] (mV/s)
Lifepak 12 (2.5-25 Hz)*	559		558	
AMSA		6.67 [4.28, 9.07] (mV-Hz)		5.02 [3.02, 7.01] (mV-Hz)
MS		3.62 [2.15, 5.10] (mV/s)		2.77 [1.50, 4.03] (mV/s)
Lifepak 15 (2.5-25 Hz)*	235		138	
AMSA		6.98 [5.19, 8.76] (mV-Hz)		5.89 [4.04, 7.75] (mV-Hz)
MS		3.91 [2.70, 5.12] (mV/s)		3.39 [2.23, 4.54] (mV/s)
Total	1372		1283	
AMSA		6.27 [4.03, 8.51] (mV-Hz)		5.02 [3.06, 6.99] (mV-Hz)
MS		3.51 [2.17, 4.86] (mV/s)		2.80 [1.56, 4.05] (mV/s)

Supplementary Appendix B: Change in Waveform Measure AUC

Increases in waveform measure prognostic performance due to inclusion of *Prior ROR* in logistic models, on test data with and without CPR, are shown below for all shocks combined and individually (Figure S1).

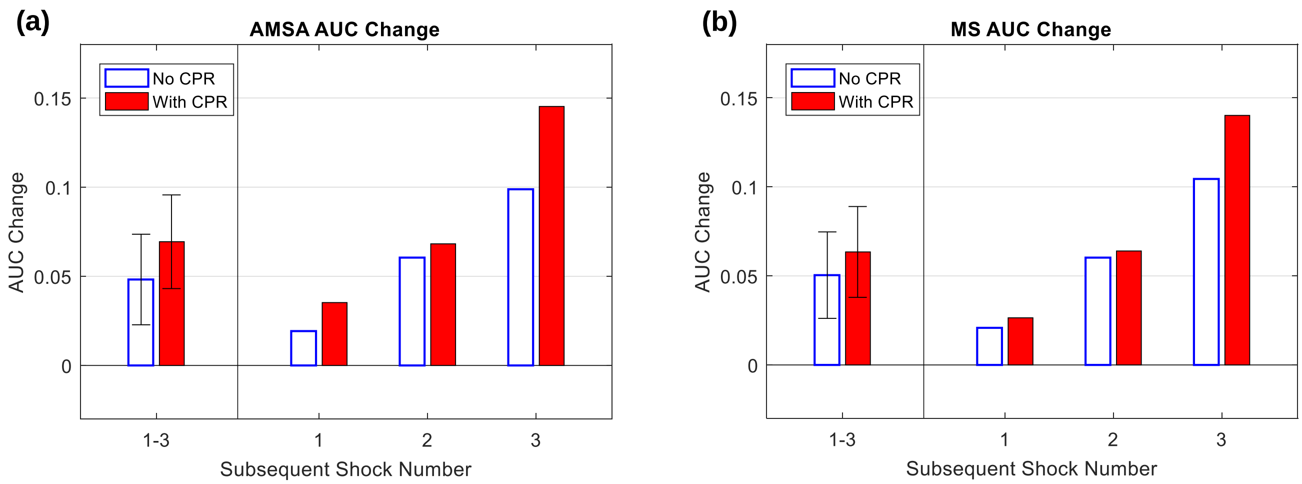


Figure S1: AUC increases due to inclusion of Prior ROR in test group (N=484 patients, n=954 CPR segments from shocks 2-4, n=884 CPR-free segments from shocks 2-4). AUC increase due to including Prior ROR is shown for individual shocks and all shocks combined for (a) AMSA prediction of ROR and (b) MS prediction of ROR. AUC increase due to Prior ROR inclusion is greater as shock sequence progresses. Error bars on combined shocks represent 95% CI for AUC change. (AMSA = Amplitude Spectrum Area; AUC = Area under the Receiver Operating Characteristic Curve; CI = Confidence Interval; CPR = Cardiopulmonary Resuscitation; MS = Median Slope; ROR = Return of Organized Rhythm)

Supplementary Appendix C: K-Fold Cross-Validation

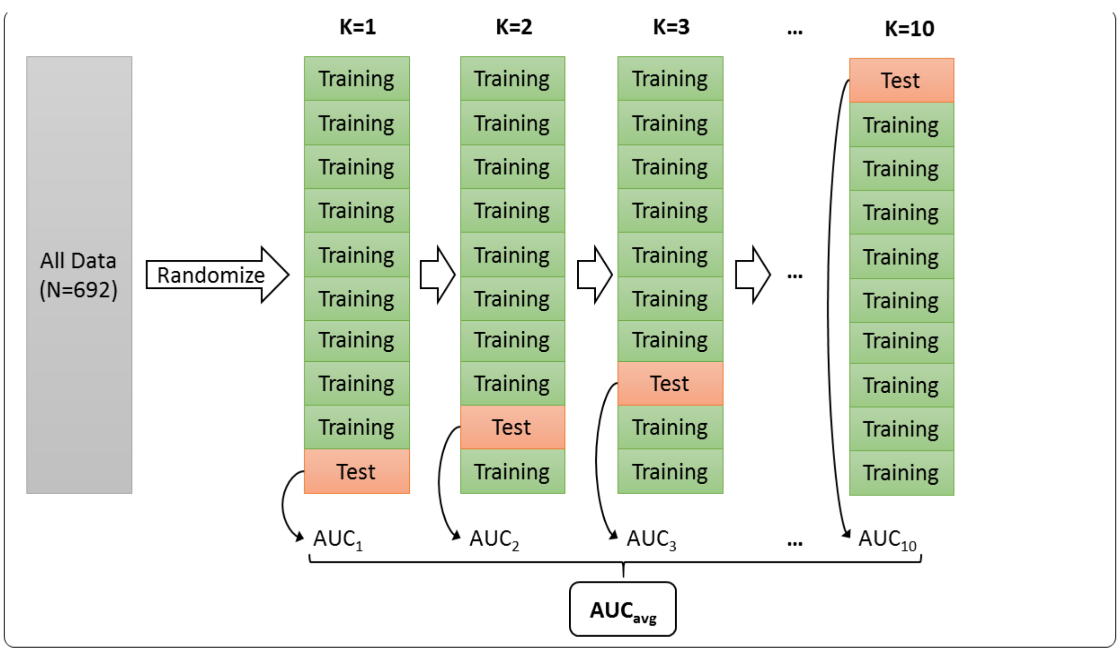
Supplementary Appendix C: Introduction

The random 30% training and 70% test split used in the primary study could produce variable results depending on which patients were assigned to the training or test groups and what percentage of the data was included in each set. For instance, a patient with ECG segments exhibiting relatively extreme CPR artifact could affect training or test results. Or, if the training set were too small, classification models might not sufficiently train and thus underestimate the benefit of including prior return of rhythm (*Prior ROR*) in combination with waveform measures. Hence, we sought to confirm the AUC values reported in the primary study using k-fold cross-validation across all samples.

Supplementary Appendix C: Methods

As an alternative to randomly splitting the patients into training and test datasets with a single division, k-fold cross validation was used across *all* ECG segments to compare the performance of waveform measures in combination with *Prior ROR* versus waveform measures alone. K-fold cross-validation is typically used to provide an estimate of classification model performance when dataset size is modest. K-fold cross-validation allows use of as large a training set as possible to allow optimal model training, while also ensuring each sample is used once in the validation results and no difficult samples are excluded.

To perform cross-validation, all ECG segments from the first three subsequent shocks were separated into k=10 exclusive, random folds of equal size for both the CPR and non-CPR datasets. Individual VF segments were divided randomly regardless of which patient they were collected from. For each of the k=10 iterations, a logistic regression model using AMSA with *Prior ROR* was trained on nine of the ten folds (Figure S2). AUCs for the logistic AMSA with *Prior ROR* model and for AMSA alone were computed on the remaining validation fold. This process was repeated for each of the ten folds, resulting in ten validation AUC values for AMSA with prior ROR and for AMSA alone. Point estimates and standard deviations of mean validation AUCs for AMSA and AMSA with prior ROR, for both without-CPR and with-CPR clips, were then compared. The procedure was then repeated for MS instead of AMSA.



<u>Figure S2</u>: Illustration of K-fold cross validation calculation of mean AUC. Data (1372 clips with CPR, or 1283 clips without CPR, from the N=692 study group patients) was randomized into k=10 equally-sized folds. Each fold was used once as a test set after training on all other folds. The AUCs for the test folds was averaged to produce the mean AUC estimate. (AUC = Area under the Receiver Operating Characteristic Curve)

Supplementary Appendix C: Results

For the first three subsequent shocks with CPR (n=1372), mean cross-validation AUC was 0.67 for AMSA alone and 0.73 for AMSA with *Prior ROR*. For MS during CPR, mean AUC was 0.68 for MS alone and 0.74 for MS with *Prior ROR*. For the first three subsequent shocks without CPR (n=1283), mean cross-validation AUC was 0.71 for AMSA and 0.75 for AMSA with *Prior ROR*. For MS without CPR, mean AUC was 0.71 for MS and 0.76 for MS with *Prior ROR* (Figure S3, Table S3).

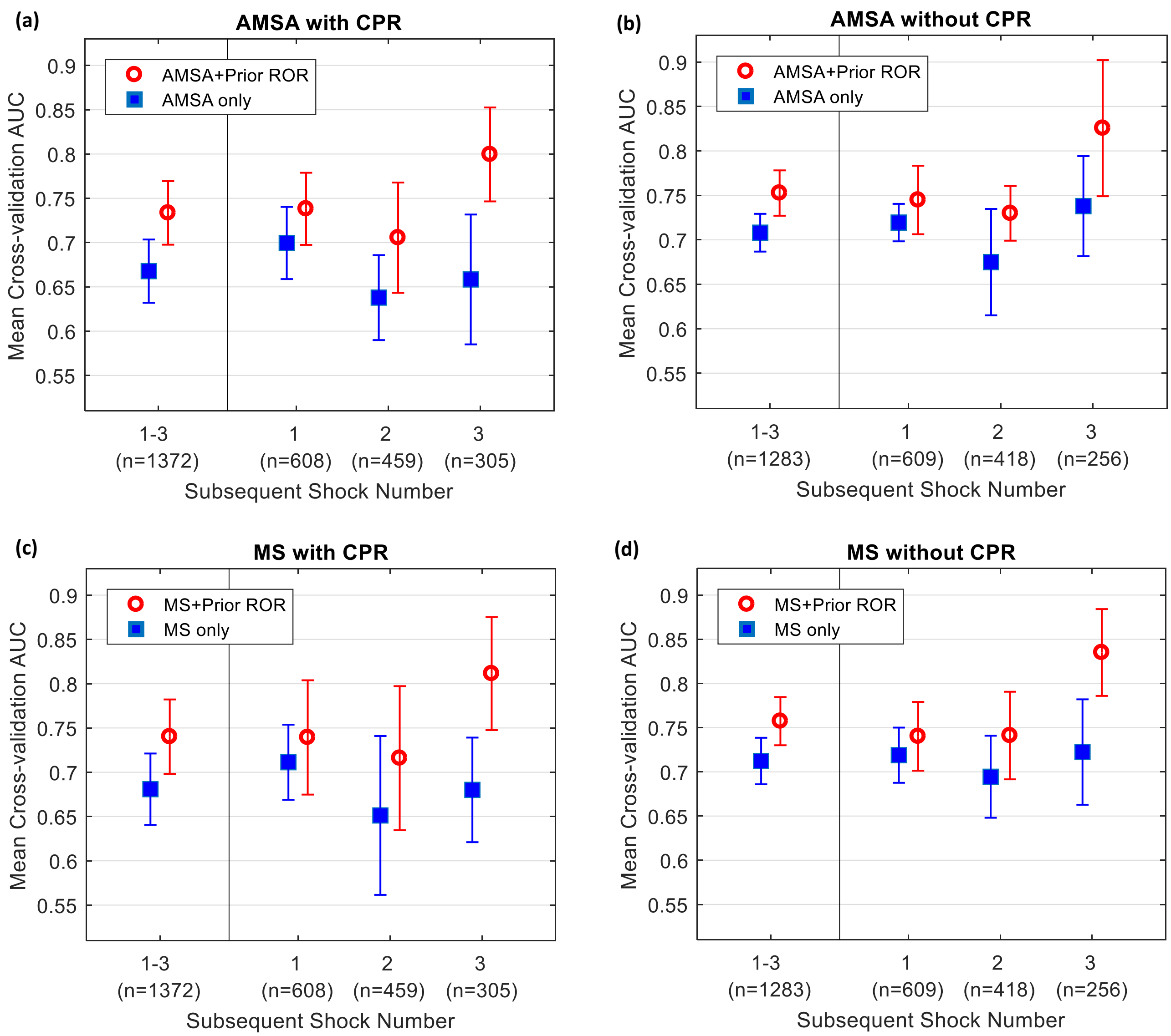


Figure S3: Mean 10-fold cross-validation AUCs for waveform measures with and without Prior ROR using all available subsequent shocks, for (a) AMSA during CPR, (b) AMSA without CPR, (c) MS during CPR, and (d) MS without CPR. Error bars represent standard deviation of the 10 cross-validation AUCs. (AMSA = Amplitude Spectrum Area; AUC = Area under Receiver Operating Characteristic Curve; CPR = Cardiopulmonary Resuscitation; MS = Median Slope; ROR = Return of Organized Rhythm)

Table S3: Mean cross-validation AUC for AMSA, MS, AMSA with *Prior ROR*, and MS with *Prior ROR* for the first three subsequent shocks following initial shock. (AMSA = Amplitude Spectrum Area; AUC = Area under Receiver Operating Characteristic Curve; CPR = Cardiopulmonary Resuscitation; MS = Median Slope; ROR = Return of Organized Rhythm)

	With CPR (n=1372)				Without CPR (n=1283)			
	Subs. Shocks 1-3	Subs. Shock 1	Subs. Shock 2	Subs. Shock 3	Subs. Shocks 1-3	Subs. Shock 1	Subs. Shock 2	Subs. Shock 3
AMSA AUC	0.67	0.70	0.64	0.66	0.71	0.72	0.68	0.74
AMSA + Prior ROR AUC	0.73	0.74	0.71	0.80	0.75	0.74	0.73	0.83
MS AUC	0.68	0.71	0.65	0.68	0.71	0.72	0.69	0.72
MS + PRIOR ROR AUC	0.74	0.74	0.72	0.81	0.76	0.74	0.74	0.84

For all subsequent shocks, combining *Prior ROR* with AMSA improved mean AUC for prediction of shock success by 0.06 during CPR and by 0.04 without CPR. Similarly, combining *Prior ROR* with MS improved mean MS AUC by 0.06 during CPR and by 0.05 without CPR (Figure S4).

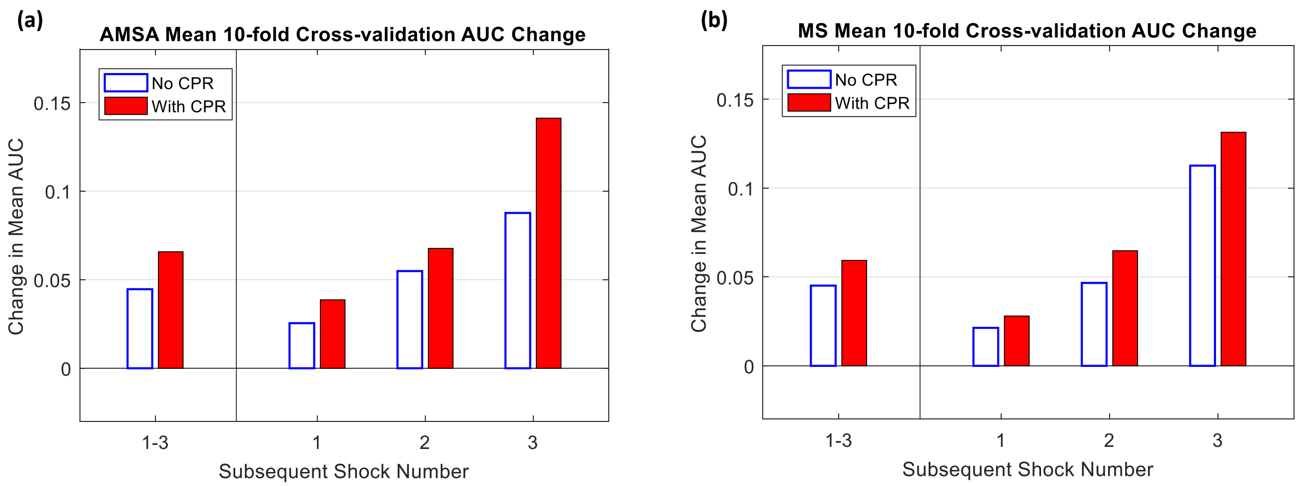


Figure S4: Increase in mean cross-validation AUC due to including *Prior ROR* in logistic model for (a) AMSA and (b) MS. (AMSA = Amplitude Spectrum Area; AUC = Area under Receiver Operating Characteristic Curve; CPR = Cardiopulmonary Resuscitation; MS = Median Slope; ROR = Return of Organized Rhythm)

Supplementary Appendix C: Discussion

Results with k-fold cross-validation confirm the primary study results. Inclusion of the prior shock's return of organized rhythm status improved the ability of AMSA and MS to predict shock success both during and without CPR. Mean cross-validation AUC for AMSA improved by approximately 0.06 for clips during CPR and 0.04 for clips without CPR, which is similar to the primary study results. Similarly, mean MS AUC increased by 0.06 during CPR and 0.05 without CPR, which is also consistent with the primary study results.

Mean k-fold cross-validation may provide a more stable estimate of model performance on limited datasets by ensuring that each and every sample, including any potentially difficult outliers, is used exactly once for validation (i.e., no difficult samples are left out of the test group). Model overfitting is also reduced due to larger training set size for each iteration. These results confirm that dataset size did not likely confound the primary study results, and that randomization of training and test clips did not affect the difficulty or generalizability of the test set.

Supplementary Appendix C: Limitations

The DeLong method used to compute p-values comparing AUCs in the primary study does not apply to mean cross-validation AUC values computed in this Appendix. The number of folds was arbitrarily selected and may affect AUC due to the fact that a higher number of validation folds, having a small number of samples in each fold, may produce unusually high or low AUC due to sample randomization and may skew the mean AUC. In the cross-validation, individual patients were not randomized (as in the primary study); this may allow clips from the same patient to be training and validation groups simultaneously.

Supplementary Appendix C: Conclusions

K-fold cross-validation results generally confirm the primary study results. However, a separate validation dataset should be evaluated for a truly independent confirmation of the primary study results.

Supplementary Appendix D: ECG Filtering and Effect of CPR

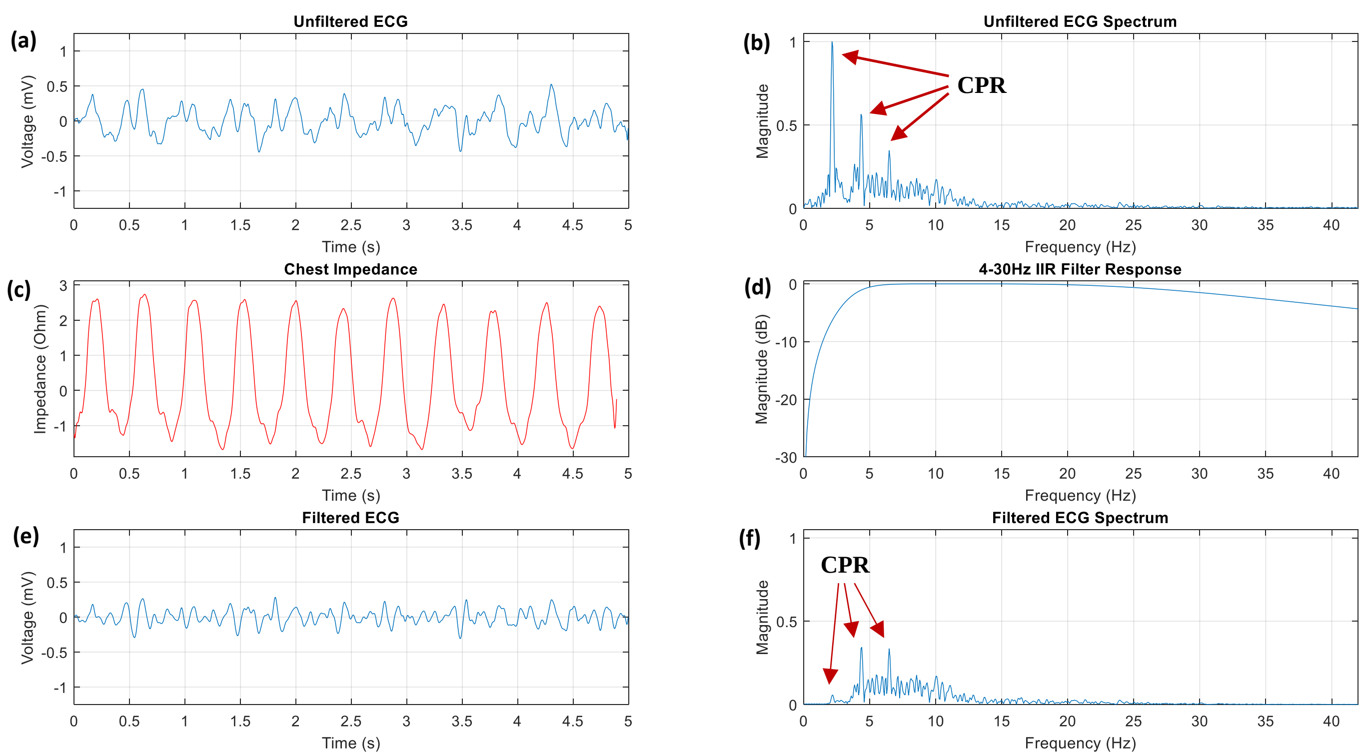
Supplementary Appendix D: Introduction

Appendix D examines the effect of the bandpass filter employed in the primary study to filter the ventricular fibrillation (VF) electrocardiogram (ECG) segments, as well as the effect of CPR artifact. Amplitude Spectrum Area (AMSA) and Median Slope (MS) waveform measures, calculated both with and without cardiopulmonary resuscitation (CPR) artifact, are compared. The objective of Appendix D is to assess the benefit of the filter on waveform measure calculation with and without CPR, and to also examine the reduction in AMSA and MS performance caused by CPR artifact.

Supplementary Appendix D: Methods

In the primary study, VF ECG segments were collected with and without CPR prior to the first three subsequent shocks from out-of-hospital primary VF patients from 2005-2014 in King County WA. Initial shocks were ineligible for the primary study because the study involved prior shock response. However, to examine the effect of the ECG bandpass filter on AMSA and MS alone in Appendix D, the eligibility criteria for VF segments is expanded to include VF collected prior to *initial* shocks in addition to subsequent shocks. Therefore, ECG segments that were collected prior to *both* the initial shock and the first three subsequent shocks when available, with and without CPR chest compression artifact, are included in this Appendix. This criteria allows for an increased number of VF samples and a more accurate assessment of filter effect on waveform measures.

After subtracting the mean from each the ECG segments, a 4^{th}-order 4-30 Hz Butterworth filter was used to filter all ECGs to mitigate the effect of chest compressions. The filter was implemented with a forward-backward implementation, which corrects phase distortion and squares the filter magnitude response. An example with CPR is shown in Figure S5 below.



<u>Figure S5:</u> (a) Unfiltered VF ECG segment during collected during CPR, with compression artifact, (b) Normalized magnitude spectrum of unfiltered ECG segment showing chest compression artifact at approximately 2.3 Hz, with harmonics at approximately 4.6 Hz and 6.9 Hz, (c) Chest impedance signal (shifted for visibility) confirms presence of CPR, (d) 4th-order 4-30 Hz Butterworth bandpass filter magnitude response, (e) Filtered ECG segment with reduced CPR artifact, (f) Spectrum of ECG showing removal of fundamental CPR frequency and a reduction in the first CPR harmonic. (CPR = Cardiopulmonary Resuscitation; ECG = Electrocardiogram; IIR = Infinite Impulse Response)

AMSA and MS were calculated using all available VF segments collected from primary VF patients within the study period, *including* segments prior to initial shock. Area under the receiver operating characteristic curve (AUC) for prediction of shock success was computed both with and without CPR for filtered and unfiltered AMSA and MS. As in the primary study, DeLong's nonparametric method for paired receiver operating characteristic curves was used to compare filtered versus unfiltered AUCs, and Robin's extension of DeLong's method for unpaired curves were used to compare AUCs with and without CPR.

Supplementary Appendix D: Results

Expanding eligibility to include patients with only an initial shock available added an additional 376 patients to the Appendix D group compared to the original primary study group, resulting in a total of 1068 patients with usable VF segments prior to shock. Demographics for the expanded Appendix D study group had slightly improved outcome compared to the original study group but were otherwise similar (Table S4).

Table S4: Demographics of Appendix D group (includes patients with just an initial shock) versus all VF and original study group. (CPR = Cardiopulmonary Resuscitation; EMS = Emergency Medical Services; IQR = Interquartile Range; ROSC = Return of Spontaneous Circulation)

	All Primary VF Cases 2005-2014 (N=1927)	Appendix D Cases (Including Initial Shocks) (N=1068)	Primary Study Group (Cases with >1 Shock) (N=692)
Female, n(%)	454(23.6)	257(24.1)	164(23.7)
Age, median (IQR)	62(52, 73)	62(52, 73)	61(52, 73)
Cardiac etiology, n(%)	1742(90.4)	968(90.6)	643(92.9)
Location, n(%)			
Home	1208(62.7)	676(63.3)	453(65.5)
Public	627(32.5)	343(32.1)	219(31.6)
Nursing Home	92(4.8)	47(4.6)	20(2.9)
Arrest before EMS arrival, n(%)	1713(88.9)	1005(94.1)	655(94.7)
Witnessed, n(%)	1482(76.9)	788(73.8)	508(73.4)
Bystander CPR, n(%)	1249(64.8)	735(68.8)	461(66.6)
EMS Response time (minutes), median (IQR)	5(4, 6.6)	5(4, 6.2)	5(4, 6)
Total shocks, median (IQR)	3(1, 6)	3(2, 6)	5(3, 7)
ROSC, n(%)	1281(66.5)	708(66.3)	435(62.9)
Died in Field	498(25.8)	277(25.9)	195(28.2)
Admit to hospital	1254(65.1)	705(66.0)	437(63.2)
Survive to hospital discharge, n(%)	818(42.4)	456(42.7)	274(39.6)

For the 1068 patients in the Appendix D study group, 1884 VF segments were collected during CPR and 2256 VF segments were collected without CPR prior to either the initial shock or first three subsequent shocks. During CPR, filtering the ECG improved AMSA AUC from 0.654 to 0.666 ($p<0.001$), and improved MS AUC from 0.0657 to 0.0685 ($p<0.001$) (Figure S6). Without CPR, filtering insignificantly reduced AMSA AUC from 0.723 to 0.722 ($p=0.378$) and insignificantly reduced MS AUC from 0.0733 to 0.728 ($p=0.063$). AUC values for AMSA and MS prediction of shock success were significantly lower during CPR than without CPR, both for unfiltered data (AMSA $p<0.001$, MS $p<0.001$) and filtered data (AMSA $p=0.001$, MS $p=0.011$) (Figure S6).

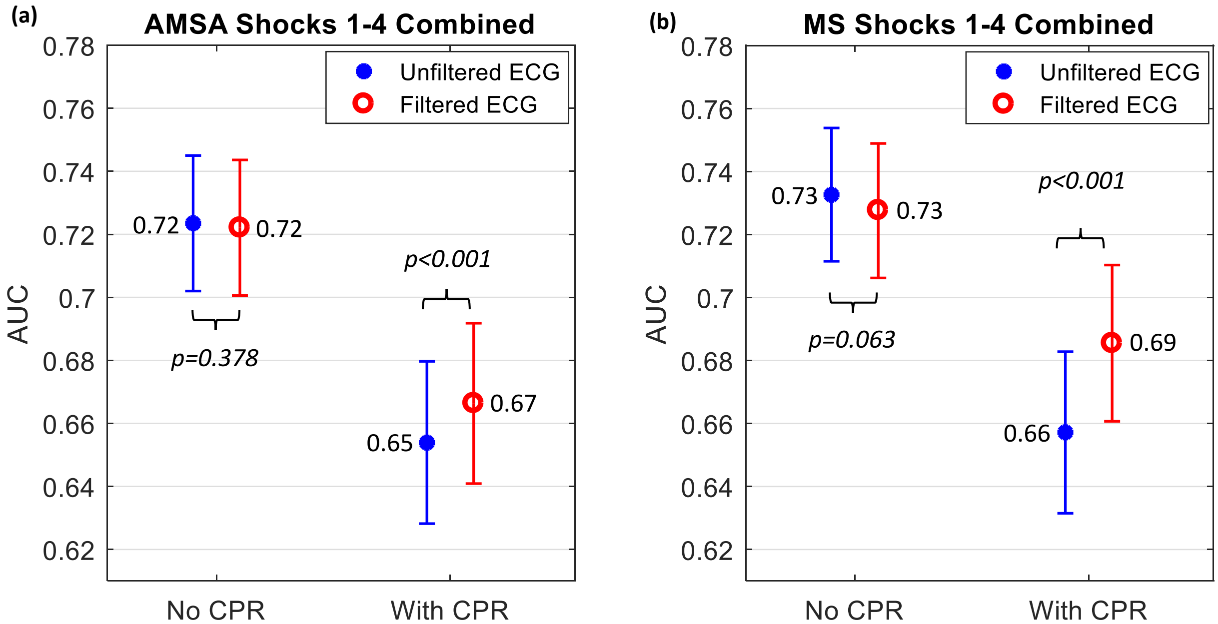


Figure S6: AMSA (a) and MS (b) AUCs with 95% CI comparison for filtered and unfiltered data for initial shocks and the first three subsequent shocks combined, from 1068 patients, with CPR (n=1884 shocks) and without CPR (n=2256 shocks). CPR artifact significantly reduced waveform measure AUC, and filtering significantly improved AUC during CPR. Note that while the bootstrap-estimated 95% confidence intervals for individual AUCs have a large overlap, the paired-sample assumption in DeLong's method allows for greater discriminatory power between two receiver operating characteristic curves calculated from the same datasets (in this case, the filtered versus unfiltered comparisons), and shows a small but significant difference between them. (AMSA = Amplitude Spectrum Area; AUC = Area under Receiver Operating Characteristic Curve; CPR = Cardiopulmonary Resuscitation; ECG = Electrocardiogram; MS = Median Slope; ROR = Return of Organized Rhythm)

Supplementary Appendix D: Discussion

AMSA and MS both inherently act as high-pass filters. AMSA, as calculated here, ignores all frequencies below 1 Hz, and multiplies each frequency magnitude between 1-26 Hz its frequency value before summing all magnitudes to produce the AMSA. Therefore, the frequency response prior to summing is a high-pass ramp. For instance, frequency content at 20 Hz has twice the contribution towards the final AMSA than frequency content at 10 Hz. Similarly, MS is the median absolute value of the slope, or pointwise difference, in the signal. The pointwise difference (derivative) operation in MS affects the frequency content of the signal by also acting as a high-pass ramp in the frequency domain.

Therefore, compared to other waveform measures, such as the median amplitude or centroid frequency, AMSA and MS may inherently benefit less from bandpass filtering the ECG, since their computation already emphasizes high frequency content by design. Other waveform measures such as the integral of the time-domain signal would likely be much more negatively affected by chest compression artifact, and would thus benefit more from ECG bandpass filtering. Even so, given our large sample size, our results show that filtering increases AUC significantly for AMSA and MS during CPR, removing much of the fundamental CPR frequency and some of the first harmonics of CPR. Filtering also reduces waveform measure AUC slightly (albeit insignificantly) without CPR by removing some of the original VF signal, due to the filter high-pass cutoff being set to 4 Hz, which slightly overlaps VF spectral content. Because of overlap between VF frequencies and CPR harmonics, choosing higher or lower cutoffs is a trade-off between improving performance either with or without compression artifact. Thus, ideally in future study, different filters would be used on CPR-artifacted signals than those used on CPR-free signals.

Supplementary Appendix D: Conclusions

Conventional (i.e., Butterworth) bandpass filtering improved waveform measure AUC during continuous CPR, and did not significantly reduce AUC without CPR. Even after filtering, however, AMSA and MS AUCs are significantly lower during CPR than without CPR.