Critical Care

Heterogeneity of treatment effect by baseline risk of mortality in critically ill patients: re-analysis of three recent sepsis and ARDS randomised controlled trials

Shalini Santhakumaran[1], Anthony Gordon[2], A. Toby Prevost[1], Cecilia O'Kane[3], Daniel F. McAuley[3,4] and Manu Shankar-Hari[5,6]*

Abstract

Background: Randomised controlled trials (RCTs) enrolling patients with sepsis or acute respiratory distress syndrome (ARDS) generate heterogeneous trial populations. Non-random variation in the treatment effect of an intervention due to differences in the baseline risk of death between patients in a population represents one form of heterogeneity of treatment effect (HTE). We assessed whether HTE in two sepsis and one ARDS RCTs could explain indeterminate trial results and inform future trial design.

Methods: We assessed HTE for vasopressin, hydrocortisone and levosimendan in sepsis and simvastatin in ARDS patients, on 28-day mortality, using the total Acute Physiology And Chronic Health Evaluation II (APACHE II) score as the baseline risk measurement, comparing above (high) and below (low) the median score. Secondary risk measures were the acute physiology component of APACHE II and predicted risk of mortality using the APACHE II score. HTE was quantified both in additive (difference in risk difference (RD)) and multiplicative (ratio of relative risks (RR)) scales using estimated treatment differences from a logistic regression model with treatment risk as the interaction term.

Results: The ratio of the odds of death in the highest APACHE II quartile was 4.9 to 7.4 times compared to the lowest quartile, across the three trials. We did not observe HTE for vasopressin, hydrocortisone and levosimendan in the two sepsis trials. In the HARP-2 trial, simvastatin reduced mortality in the low APACHE II group and increased mortality in the high APACHE II group (difference in RD = 0.34 (0.12, 0.55) ($p = 0.02$); ratio of RR 3.57 (1.77, 7.17) ($p < 0.001$). The HTE patterns were inconsistent across the secondary risk measures. The sensitivity analyses of HTE effects for vasopressin, hydrocortisone and levosimendan were consistent with the main analyses and attenuated for simvastatin.

Conclusions: We assessed HTE in three recent ICU RCTs, using multivariable baseline risk of death models. There was considerable within-trial variation in the baseline risk of death. We observed potential HTE for simvastatin in ARDS, but no evidence of HTE for vasopressin, hydrocortisone or levosimendan in the two sepsis trials. Our findings could be explained either by true lack of HTE (no benefit of vasopressin, hydrocortisone or levosimendan vs comparator for any patient subgroups) or by lack of power to detect HTE. Our results require validation using similar trial databases.

Keywords: Sepsis, acute respiratory distress syndrome, Models, statistical, Randomisation, Risk, Study design

* Correspondence: manu.shankar-hari@kcl.ac.uk
[5]Department of Intensive Care Medicine, St Thomas' Hospital, Guy's and St Thomas' NHS Foundation Trust , Westminster Bridge Road, London SE1 7EH, UK
[6]Peter Gorer Department of Immunobiology, School of Immunology & Microbial Sciences, King's College London, London SE1 9RT, UK
Full list of author information is available at the end of the article

Santhakumaran *et al. Critical Care*   (2019) 23:156

Page 2 of 9

Background

Non-random variation in the treatment effect of an intervention due to differences in the baseline risk of death between patients in a population represents one form of heterogeneity of treatment effect (HTE) [1, 2]. In critical care settings, sepsis [3] and acute respiratory distress syndrome (ARDS) [4] are acute illnesses with significant clinical and biological heterogeneity [5–8]. Therefore, it is possible that even in randomised controlled trials (RCTs) which enrol patients that meet specific sepsis or ARDS eligibility criteria, there may still be heterogeneity in the trial populations. This heterogeneity occurs both within a trial and between trials [9]. The resulting variation in risk of outcomes may result in clinically important HTE in such trial populations. This heterogeneity is one possible explanation for indeterminate results in sepsis and ARDS RCT [9, 10]. We use the term *indeterminate* to illustrate that statistically non-significant results of two-tailed tests suggest uncertainty in results, as opposed to proof of no difference between treatments, implied by the term *negative* [11].

Recently, Iwashyna and colleagues simulated RCTs using observational cohort data and highlighted that the magnitude of HTE may be such that the average benefit (or harm) from the tested treatment in critical care RCTs may not be valid for an individual patient meeting the trial eligibility criteria [10]. Therefore, exploring HTE with data from completed RCTs where the intervention showed no effect in the overall population, aside from explaining the RCT results, could also inform future trial design and trial efficiency by targeting a trial population defined by a specific baseline measure associated either with the highest treatment benefit or with treatment response (enrichment) [9, 12].

In this context, we explored the presence of HTE for vasopressin and hydrocortisone in the VANISH trial [13], for levosimendan in the LeoPARDS trial [14] and for simvastatin in the HARP-2 trial [15]. We hypothesised that the individual patient's baseline risk of death modifies the direction and magnitude of the treatment effects of vasopressin [13], hydrocortisone [13], levosimendan [14] and simvastatin [15] within these RCTs. A number of recent studies support our hypothesis. The treatment effect of vasopressin differed with severity of septic shock in a previous RCT [16]. The treatment effect of hydrocortisone differed between trials [17], with potential benefit seen in trials with higher control group mortality [18–20]. The treatment effect of simvastatin differed between ARDS sub-phenotypes [21] and potentially with illness severity in critically ill patients [22].

Our overall aim was to assess whether the individual patient's baseline risk of death modifies the treatment effect of an intervention (HTE). The Acute Physiology And Chronic Health Evaluation II (APACHE II) model has been proposed as a potential model for HTE evaluation [10, 23, 24]. We assessed HTE using the APACHE II score [24] as the primary measure of baseline risk, and two secondary measures based on the APACHE II model: the APACHE II physiology score (APS-APII), and the APACHE II calculated risk of death as originally proposed by Knauss and colleagues (R) [24]. The rationale for using the APS-APII was that the total APACHE II score consists of non-modifiable risk of death from age and comorbidity, but the physiological derangement most likely mediates the treatment effect to outcome relationship [25]. Variation in the absolute risk difference may occur even if the relative effect of the treatment is the same, whilst the relative risk associated with the treatment may also vary. Therefore, we examined absolute and relative measures of heterogeneity. We also investigated whether any HTE could be driven by adverse events, as low-risk patients may have similar exposure to treatment-related harms to the high-risk patients, but not to the benefits, resulting in a net harm signal [10]. Furthermore, irrespective of whether the treatment effects of interventions varied or remained constant over the range of baseline risk, HTE may manifest due to differences in treatment-related adverse events over the range of baseline risk.

Methods

Study approvals and RCT datasets

We obtained ethics approval for this study (18/LO/1079). VANISH was a 2×2 factorial, double-blind, RCT in adult patients with sepsis who required vasopressors, in 18 general adult intensive care units (ICUs) in the United Kingdom (UK) [13]. In the VANISH trial [13], patients were randomly allocated to vasopressin and hydrocortisone ($n = 101$), vasopressin and placebo ($n = 104$), norepinephrine and hydrocortisone ($n = 101$) or norepinephrine and placebo ($n = 103$). Patients only received the second study drug (hydrocortisone/placebo) if the maximum infusion of the first study drug (vasopressin/norepinephrine) had been reached. Therefore, in the hydrocortisone analysis, only participants who received the study drug were included (hydrocortisone $n = 148$, placebo $n = 148$); all remaining analyses are intention-to-treat. The 28-day mortality was 63/204 (30.9%) of patients in the vasopressin group and 56/204 (27.5%) patients in the norepinephrine group (difference = 3.4% [95% CI, – 5.4–12.3%]) [13]. LeoPARDS was a two-arm parallel group, double-blind, placebo-controlled RCT in adult patients with sepsis who required vasopressors, in 34 ICUs in the UK [13]. In LeoPARDS trial [14], patients were randomised to receive either levosimendan ($n = 258$) or placebo ($n = 257$) over 24 h in addition to standard care. The 28-day mortality was 89/258 (34.5%) in the levosimendan group and 79/256 (30.9%) in the placebo group (difference = 3.6% [95% CI, – 4.5–11.7%]) [14]. HARP-2 was a two-arm parallel group, double-blind, placebo-controlled RCT in adult patients within 48 h after

Santhakumaran et al. Critical Care   (2019) 23:156

Page 3 of 9

the onset of ARDS in 40 ICUs in the UK and Ireland [15]. In the HARP-2 trial [15], patients were randomised to receive either once-daily simvastatin or identical placebo tablets enterally for up to 28 days. The 28-day mortality was 57/259 (22.0%) in the simvastatin group and 75/280 (26.8%) in the placebo group (risk ratio = 0.8 [95% CI, 0.6 to 1.1]) [15].]

Statistics

The primary analysis examined HTE for 28-day mortality with APACHE II score as the measure of baseline risk, comparing treatment effect in patients above (high) and below (low) the overall median score of 25. As secondary analyses, we examined two other baseline risk measures, APS-APII and R. Distributions of these baseline risk measures and mortality were described with histograms, and the discriminatory performance was assessed using the area under the receiver operating characteristic curve (AUC). We estimated the extreme quartile odds ratio (EQuOR, the ratio of the odds of death in the highest vs. lowest quartile for risk) as an estimate of how the risk of death varies between patients in the same trial [26]. Forest plots illustrated the absolute risk difference (RD) and relative risk (RR) for 28-day mortality by treatment group comparing high and low APACHE II groups. HTE was quantified on both the absolute and relative scales via additive and multiplicative interactions respectively. The difference in the RD and associated 95% confidence interval (CI) was estimated assuming a linear model for the probability of death, with treatment, a binary indicator for APACHE II subgroup, and the interaction between them as covariates, using robust standard errors. The ratio of the RR and 95% CI was estimated assuming a log-binomial model with the same covariates. We then investigated heterogeneity of harms using forest plots by APACHE II subgroup similar to the primary analysis. Interactions were not estimated for heterogeneity of harms due to the low number of events. For the HARP-2 trial, only the primary baseline risk measure of the total APACHE II score was available.

Sensitivity analyses

Four sensitivity analyses for the main baseline risk measure (APACHE II score) were performed. First, we used hospital mortality as the outcome instead of at 28 days, as APACHE II score was originally devised as a prediction tool for hospital mortality. Second, we investigated the potential impact of missing data on the results. In the VANISH trial, there were 47 patients who had at least one element of the acute physiology score missing, and 61 patients in the LeoPARDS trial. In the main analysis, normal scores were assumed for these elements, as for the main trial [13, 14]. In the HARP-2 trial, 66 patients had missing total APACHE II scores and were omitted from the main analysis but displayed in the forest plot. Missingness occurred

pre-randomisation and hence is independent of the treatment effect but may affect the precision of the results. In the sensitivity analysis, we assumed patients with missing data were (i) equally likely to be in the high-risk group as those with complete data, (ii) 10% more likely and (iii) 10% less likely. APACHE II category was imputed 20 times under these assumptions, and the difference in RD and ratio of RR computed as for the main analysis, combining results across imputations using Rubin's rules. A third sensitivity analysis was performed by recalibrating the APACHE II risk prediction using the whole RCT cohort, as internally developed risk models using both treatment arms are preferred to models developed on the control population, as highlighted by Burke et al. [27]. A logistic regression model for 28-day mortality was constructed with the following covariates: APACHE II points from each acute physiology component, age points, chronic health points, post-emergency surgery and diagnostic category weight. The resulting prediction was used as a measure of baseline risk, assessing HTE as in the main analysis. To avoid spurious associations from categorisation of APACHE II score [28], we performed a fourth sensitivity analysis, by treating APACHE II as a continuous variable in a logistic regression model. Relative HTE was quantified by the interaction between APACHE II score and treatment, expressed as a ratio of odds ratios. Additive HTE was illustrated by plotting the estimated absolute difference in mortality between treatment groups across the range of APACHE II.

Results

Baseline risk of 28-day mortality

The 28-day mortality, between the intervention and control arms, in the VANISH, LeoPARDS and HARP-2 trials was not significantly different (Table 1). The illness severity (using the total APACHE II score) was lower in the HARP-2 trial, compared to those in the VANISH and LeoPARDS trials (Fig. 1). The EQuOR highlighted significant heterogeneity of risk of death in all three RCTs for all three risk measures.

VANISH trial HTE assessment

The 28-day mortality increased in the vasopressin and in the norepinephrine group, with increasing baseline risk measures (Fig. 1). For the primary analysis with APACHE II score as baseline risk of death measure, there was no evidence of HTE for vasopressin in either absolute terms (risk difference for low APACHE II 0.02 (− 0.09, 0.13) and high APACHE II 0.05 (− 0.08, 0.19); difference in risk difference 0.04 (− 0.14, 0.21)) or relative terms (relative risk for low APACHE II 1.09 (0.64, 1.86) and high APACHE II 1.15 (0.08, 1.64); ratio of relative risk 1.05 (0.55, 2.00)) (Fig. 2). For the secondary risk measures, the estimates of HTE for vasopressin were larger with wider CI for APS-APII (Fig. 3) and smaller in magnitude for R (Fig. 4).

Santhakumaran *et al. Critical Care*   (2019) 23:156

Page 4 of 9

Table 1 Trial level summary characteristics

	VANISH Trial		LeoPARDS Trial		HARP-2 Trial	
	Vasopressin	**Noradrenaline**	**Levosimendan**	**Placebo**	**Simvastatin**	**Placebo**
28-day mortality, n/N (%)	63/204 (31)	56/204 (27)	89/258 (35)	79/256 (31)	57/259 (22)	75/279 (27)
Related* AE, n/N (%)	23/205 (11)	16/204 (8)	41/258 (16)	16/257 (6)	36/259 (14)	25/279 (9)
Related* SAE, n/N (%)	13/205 (6)	10/204 (5)	13/258 (5)	2/257 (1)	3/259 (1)	4/279 (1)
APACHE II, median (IQR)	24 (19,29)	24 (19,30)	25 (21,31)	25 (21,30)	18 (14,24)	18 (14,23)
APS-APII, median (IQR)	20 (14,24)	20 (15,25)	20 (16,26)	21 (16,24)		
R, median (IQR)	0.41 (0.24,0.63)	0.42 (0.25,0.66)	0.56(0.36,0.72)	0.53 (0.39,0.70)		
	Hydrocortisone	**Placebo**				
28-day mortality, n/N (%)	52/147 (35)	47/148 (32)				
Related AE, n/N (%)	18/148 (12)	18/148 (12)				
Related SAE n/N (%)	11/148 (7)	12/148 (8)				
APACHE II, median (IQR)	25 (19,32)	25 (20,30)				
APS, median (IQR)	21 (15,26)	20 (16,25)				
R, median (IQR)	0.48 (0.25,0.69)	0.44 (0.28,0.67)				
AUC APACHE II	0.67 (0.62,0.73)		0.71 (0.66,0.75)		0.68 (0.62,0.73)	
AUC APS	0.62 (0.56, 0.68)		0.67 (0.62,0.72)			
AUC R	0.67 (0.62,0.73)		0.67 (0.62,0.72)			
EQuOR APACHE II	4.85 (2.49 to 9.46)		7.35 (4.09 to 13.20)		5.92 (2.99 to 11.73)	
EQuOR APS	3.58 (1.88 to 6.83)		5.39 (3.06 to 9.51)			
EQuOR R	5.66 (2.83 to 11.31)		4.64 (2.63 to 8.17)			

Shaded regions in the HARP-2 trial represent lack of raw data to derive APS-APII score or *R*
IQR interquartile range, *AUC* area under the receiver operating characteristic curve, *EQuOR* extreme quartile odds ratio, *(S)AE* (serious) adverse events, *APACHE II* Acute Physiology And Chronic Health Evaluation II, *APS-APII* Acute Physiology Score from APACHE II, *R* risk of death calculated from APACHE II

For the primary analysis with APACHE II score as baseline risk of death measure, there was no evidence of HTE for hydrocortisone in either absolute terms (risk difference for low APACHE II 0.02 (– 0.12, 0.17) and high APACHE II 0.06 (– 0.10, 0.21); difference in risk difference 0.03 (– 0.18, 0.25) or relative terms (relative risk for low APACHE II 1.11 (0.62, 1.99) and high APACHE II 1.15 (0.79, 1.67); ratio of relative risk 1.04 (0.52, 2.08)). For the secondary risk measures, the estimates of HTE for hydrocortisone were similar for APS-APII (Fig. 3) and larger in magnitude for *R* (Fig. 4).
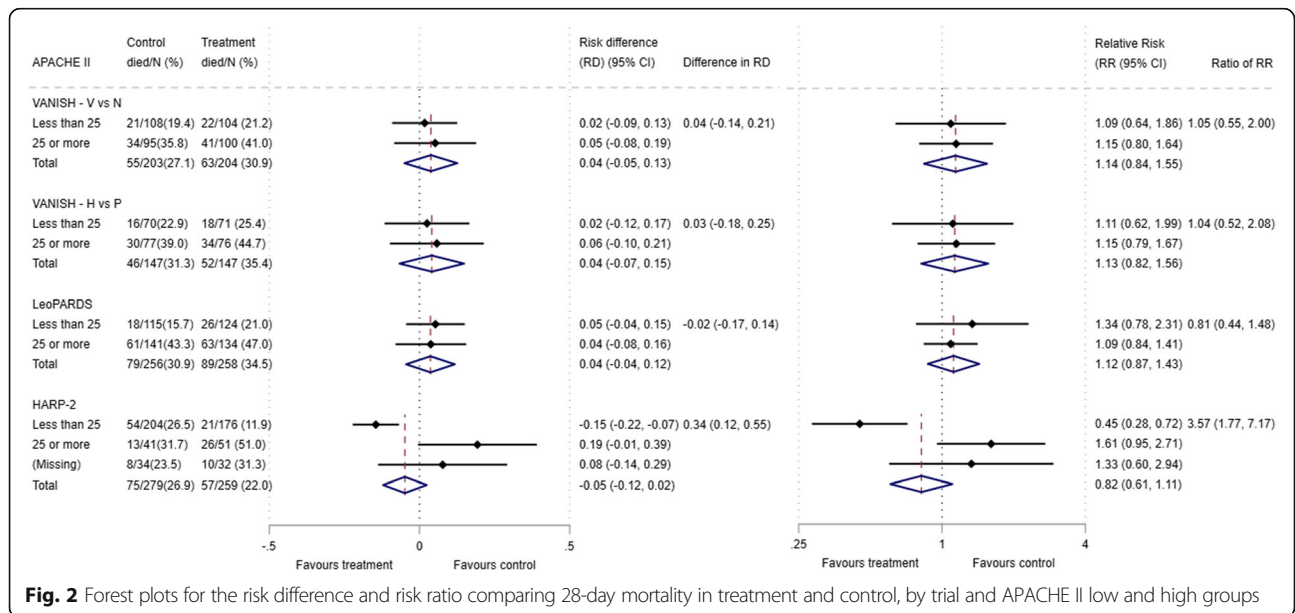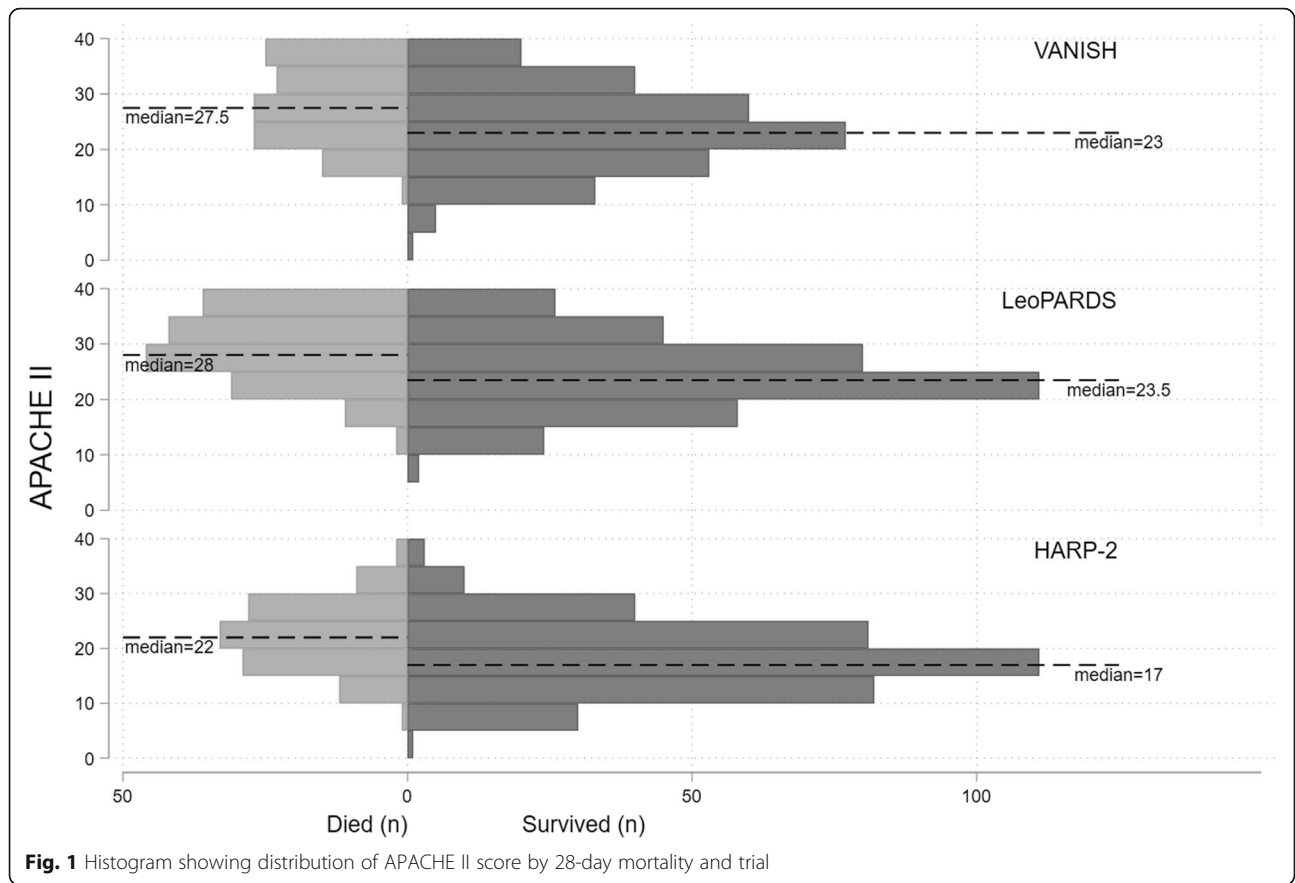
LeoPARDS trial HTE assessment
The 28-day mortality increased in the levosimendan and in the placebo group, with increasing baseline risk measures (Fig. 1). For the primary analysis with APACHE II score as baseline risk of death measure, there was no evidence of HTE for levosimendan in either absolute terms (risk difference for low APACHE II 0.05 (– 0.04, 0.15) and high APACHE II 0.04 (– 0.08, 0.16); difference in risk difference – 0.02 (– 0.17 to 0.14)) or relative terms (relative risk for low APACHE II 1.34 (0.78, 2.31) and high APACHE II 1.09 (0.84, 1.41); ratio of relative risk 0.81 (0.44 to 1.48)) (Fig. 2). For the secondary risk measures, the estimates of HTE for levosimendan were larger for APS-APII (Fig. 3) and in the opposite direction for *R* (Fig. 4).

HARP-2 trial HTE assessment
The 28-day mortality increased in the simvastatin group and in the standard care group, with increasing baseline risk measures (Fig. 1). For the primary analysis with APACHE II score as baseline risk of death measure, we observed HTE for simvastatin in absolute terms (risk difference for low APACHE II – 0.15 (– 0.22, – 0.07) and high APACHE II 0.19 (– 0.01, 0.39); difference in risk difference 0.34 (0.12, 0.55) ($p = 0.02$)) and in relative terms (relative risk for low APACHE II 0.45 (0.28, 0.72) and high APACHE II 1.61 (0.95, 2.71); ratio of relative risk 3.57 (1.77 to 7.17)). Simvastatin reduced mortality in the low APACHE II group and increased mortality in the high APACHE II group (Fig. 2). As raw data APACHE II score data were not available, we have not reported any secondary risk measures for the HARP-2 trial.

Serious adverse events and baseline risk
We plotted the proportions of serious adverse events by low and high APACHE II groups in each trial, to explore whether the pattern of adverse event distribution could explain any HTE in mortality. In all three RCTs, both in the intervention and control trial arms, there was no pattern in serious adverse events that could explain HTE in mortality (Additional file 1: Figure S1).
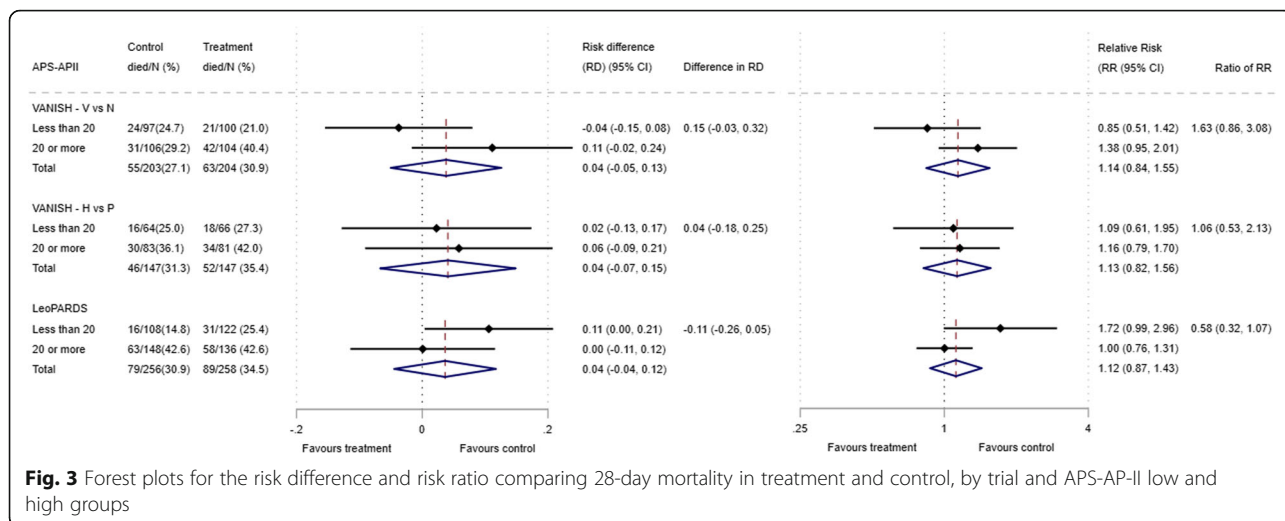
Fig. 1 Histogram showing distribution of APACHE II score by 28-day mortality and trial

Fig. 2 Forest plots for the risk difference and risk ratio comparing 28-day mortality in treatment and control, by trial and APACHE II low and high groups

Santhakumaran *et al. Critical Care*   (2019) 23:156

Page 6 of 9

Fig. 3 Forest plots for the risk difference and risk ratio comparing 28-day mortality in treatment and control, by trial and APS-AP-II low and high groups

Sensitivity analyses

Results from sensitivity analyses were consistent with the main analyses for the VANISH and LeoPARDS trials (Additional file 1: Table S1, Table S2, Figure S2, Figure S3 and Figure S4). HTE effects were attenuated in the sensitivity analyses for the HARP-2 trial under different assumptions for the missing data (e.g. ratio of relative risk was 2.86 (1.47, 5.57) when we assumed that patients with missing APACHE II data were more likely to be high risk; all other results were less attenuated Additional file 1: Table S1). Differences were also smaller when hospital mortality was used as the outcome (difference in risk difference 0.25 (0.03, 0.48); ratio of RR 2.34 (1.31, 4.18), Additional file 1: Figure S1) and when HTE was assessed across the continuous range of APACHE II score (ratio of odds ratio for a 5-point increase in APACHE II 1.33 (0.93, 1.90) Additional file 1: Table S2 and Figure S4).

Discussion

We assessed whether HTE could contribute to the indeterminate results in three recent ICU RCTs, using multivariable baseline risk of death models, which included well-established risk factors for acute mortality for sepsis and ARDS as covariates. There was considerable within-trial variation in the baseline risk of death in all three RCTs. We did not observe HTE for vasopressin, hydrocortisone and levosimendan in the two sepsis trials, though there was evidence of differential treatment effect in the HARP-2 trial for ARDS with low risk of death sub-population benefitting the most. We observed that detection of HTE in RCTs may be influenced by the baseline risk model specification, as illustrated by differences in HTE effects seen with different models reported using the LeoPARDS trial data.
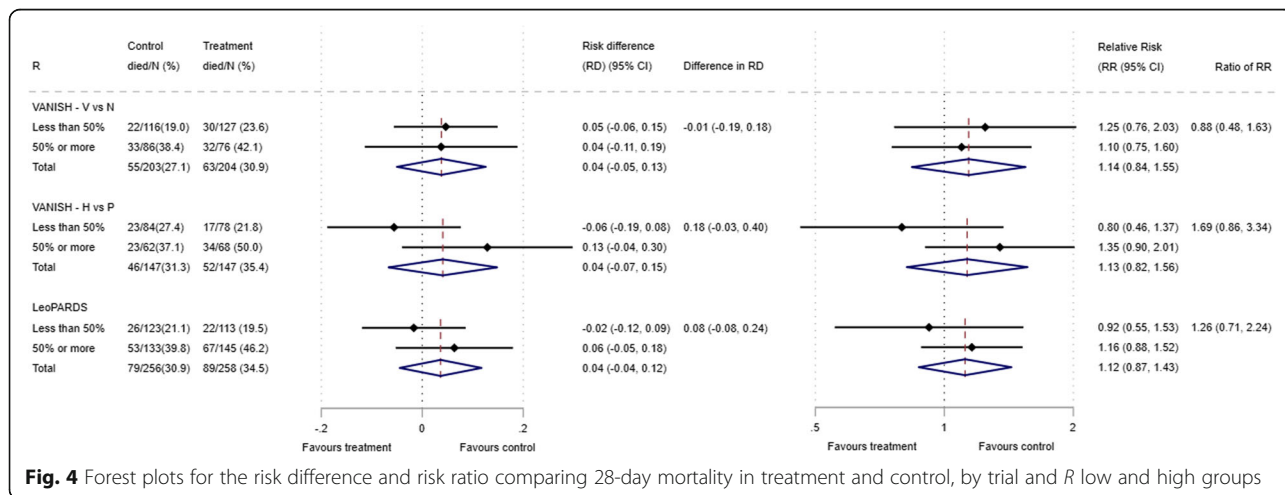

Fig. 4 Forest plots for the risk difference and risk ratio comparing 28-day mortality in treatment and control, by trial and *R* low and high groups

Santhakumaran *et al. Critical Care* (2019) 23:156

Page 7 of 9

Explanation of key findings

There are a number of possible reasons why we did not observe HTE consistently in all our analyses. All three trials we assessed have many features of explanatory trials [29], which by their design limit HTE in comparison to pragmatic trials, such as through narrower eligibility criteria, intensity of follow-up and non-mortality primary outcomes. Therefore, demonstrable HTE is less likely in these trials, though its evaluation remains important. Our findings may be true in that HTE may be less marked in sepsis and ARDS as mortality may be driven by non-modifiable risk factors such as older age and presence of comorbid conditions, alongside illness attributable risk, generating many "minimal contributory causes" of mortality [30], when compared to illnesses such as retroviral disease [31]. It could be that the effects of treatments we assessed on mortality are both small and with limited variability across baseline risk of death resulting in minimal HTE. Another explanation for not observing HTE may be that there is no true treatment effect difference between subgroups enriched on prognosis with APACHE II score. Given the sample size in the trials assessed, we may only have power to detect large interaction effects, which requires either a large treatment effect in one or more subgroups, opposing treatment effects between subgroups or differential adverse event risk between subgroups.

Comparison to published literature

A key comparison to consider is the contrasting results with RCT simulations by Iwashyna and colleagues [10]. Their simulations assumed that the trial participants' odds of 30-day mortality will be influenced by the severity of acute respiratory failure, comorbid conditions, the treatment's reduction in mortality from the primary illness and the treatment's fatal adverse effect rates. Furthermore, Iwashyna and colleagues assumed constant relative treatment effects, constant harms and mortality patterns predicted by their simulation model. We used 28-day mortality, for our primary analysis. We considered baseline risk of death as a function of acute illness severity using the total APACHE II score, which is in line with the conceptual arguments put forward by Kent and colleagues [1] that HTE emerges from the risk of outcome (28-day mortality in our study), the risk of treatment-related harm and direct treatment-effect modification. Importantly, the data in our trials do not follow the constant relative treatment effects, constant harms and therefore the mortality patterns described by Iwashyna and colleagues [10].

Recently, Semler and colleagues reported a pragmatic, cluster-randomised, multiple-crossover trial of saline versus balance crystalloids in critically ill patients, with no difference in primary outcome of major adverse kidney events within 30 days (MAKE30), a composite of in-hospital death, new receipt of renal-replacement therapy and persistent renal dysfunction [32], but reported presence of HTE for this outcome with a multivariable model specifically calibrated for this outcome [33]. In contrast, our analytic strategy ascertained whether the observed treatment effect differed by the pre-randomisation baseline risk of death multivariable model (APACHE II score), which helped us to compare multiple treatments in critically ill patients with sepsis or ARDS.

Strengths and weakness

We explored heterogeneity in absolute and in relative treatment effects, in sepsis and ARDS, for four different treatments and using three different measures of baseline risk. The primary baseline risk measure, APACHE II, is an established, validated predictor of mortality in this population. Two variations on this measure were investigated to check the consistency of the results, along with several sensitivity analyses. We used a composite risk score (APACHE II) for its superior performance for baseline risk estimation, as highlighted by Kent and colleagues [1] and a recommendation for future studies on HTE assessment [10]. There were insufficient numbers to examine HTE across smaller groups (e.g. quartiles). None of the RCTs included in this study had 28-day mortality as the primary outcome; it is possible that we were underpowered to detect HTE. The primary outcomes were not suitable for HTE analysis as they were continuous rather than binary and without an appropriate baseline measure, though the existing HTE framework could be adapted for some continuous outcomes, such as change from baseline organ dysfunction.

Implications of research

Aside from the ARDS or sepsis illness characteristics, it is likely that biological mechanisms determining differences in treatment effect will vary with the intervention tested. Therefore, using a generic physiology-based multivariable model such as APACHE II with biomarkers that provide both prognostic and predictive enrichment or intervention-specific predictive enrichment coupled may be a better approach to defining a study population. For example, an ARDS sub-population with greater inflammation and higher mortality was more likely to benefit from simvastatin [21, 34], and aside from severity of septic shock, the treatment effect of vasopressin was associated with biological differences within the trial population [16, 35]. Similarly, biomarkers derived from whole blood transcriptomics could help enrich septic shock patients for corticosteroid therapy [36–38]. It is plausible that when HTE is assessed for an intervention using data from a single trial, we are unlikely to detect it unless HTE effects are large. This generates an argument to assess HTE using trials of the similar treatment-condition combination or of the same condition

Santhakumaran *et al. Critical Care*   (2019) 23:156

Page 8 of 9

and broader group of treatments with similar enough mechanism of treatment effect, or consider intervention-specific multivariable models. As suggested by Iwashyna and colleagues, perhaps HTE assessment should form part of a priori analyses plans in future clinical trials. As HTE is about the variation in effectiveness, standardising the baseline risk measure between RCTs, including HTE assessment as a priori analyses, ensuring that the outcome used in HTE analyses is patient-centered (such as mortality) and incorporating the proposals within the Core Outcome Measures in Effectiveness Trials guidelines will enable pooling of HTE analysis across future trials [39].

Conclusions

We assessed HTE in three recent ICU RCTs, using multivariable baseline risk of death models. There was considerable within-trial variation in the baseline risk of death. We observed potential HTE for simvastatin in ARDS, but no evidence of HTE for vasopressin, hydrocortisone or levosimendan in the two sepsis trials. Our findings could be explained either by true lack of HTE (no benefit of vasopressin, hydrocortisone or levosimendan vs comparator for any patient subgroups) or by lack of power to detect HTE. Our results require validation using similar trial databases.

Additional file

Additional file 1: Table S1. Results from multiple imputation analysis; for patients with missing APACHE II, we assumed the proportion in the high-risk category (APACHE II ≥ 25) was either the same as the trial participants with complete data, 10% higher or 10% lower. **Table S2.** Treatment-risk interaction using continuous APACHE II from logistic regression analysis of 28-day mortality. **Figure S1.** Forest plots for the risk difference and risk ratio comparing related serious adverse events in treatment and control, by trial and APACHE II subgroup. **Figure S2.** Forest plots for the risk difference and risk ratio comparing hospital mortality in treatment and control, by trial and APACHE II subgroup. **Figure S3.** Forest plots for the risk difference and risk ratio comparing related serious adverse events in treatment and control, by trial and APACHE II subgroup. **Figure S4.** HTE assessment for APACHE II score as a continuous variable. Figures show the estimated treatment effect with 95% confidence interval bands from regression models for 28-day mortality including a treatment × APACHE II score interaction for **Figure S4A:** VANISH Vasopressin; **Figure S4B:** VANISH Hydrocortisone; **Figure S4C:** LeoPARDS and **Figure S4D:** HARP-2. (DOCX 577 kb)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Imperial Clinical Trials Unit, School of Public Health, Imperial College London, London W12 7RH, UK. [2]Section of Anaesthetics, Pain Medicine and Intensive Care, Imperial College London, London W2 1NY, UK. [3]Centre for Experimental Medicine, Wellcome-Wolfson Institute for Experimental Medicine, Belfast BT9 7AE, UK. [4]Regional Intensive Care Unit, Royal Victoria Hospital, Belfast BT12 6BA, UK. [5]Department of Intensive Care Medicine, St Thomas' Hospital, Guy's and St Thomas' NHS Foundation Trust , Westminster Bridge Road, London SE1 7EH, UK. [6]Peter Gorer Department of Immunobiology, School of Immunology & Microbial Sciences, King's College London, London SE1 9RT, UK.

References

1. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials. 2010;11:85.
2. Senn S. Mastering variation: variance components and personalised medicine. Stat Med. 2016;35(7):966–77.
3. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801–10.
4. Force ADT, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, Camporota L, Slutsky AS. Acute respiratory distress syndrome: the Berlin definition. JAMA. 2012;307(23):2526–33.
5. Scicluna BP, van Vught LA, Zwinderman AH, Wiewel MA, Davenport EE, Burnham KL, Nurnberg P, Schultz MJ, Horn J, Cremer OL, et al. Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. Lancet Respir Med. 2017;5(10):816–26.
6. Davenport EE, Burnham KL, Radhakrishnan J, Humburg P, Hutton P, Mills TC, Rautanen A, Gordon AC, Garrard C, Hill AV, et al. Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. Lancet Respir Med. 2016;4(4):259–71.
7. Bos LD, Schouten LR, van Vught LA, Wiewel MA, Ong DSY, Cremer O, Artigas A, Martin-Loeches I, Hoogendijk AJ, van der Poll T, et al. Identification and validation of distinct biological phenotypes in patients with acute respiratory distress syndrome by cluster analysis. Thorax. 2017;72(10):876–83.

8.   Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA, Network NA. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. Lancet Respir Med. 2014;2(8):611–20.

9.   Shankar-Hari M, Rubenfeld GD. The use of enrichment to reduce statistically indeterminate or negative trials in critical care. Anaesthesia. 2017;72(5):560–5.

10. Iwashyna TJ, Burke JF, Sussman JB, Prescott HC, Hayward RA, Angus DC. Implications of heterogeneity of treatment effect for reporting and analysis of randomized trials in critical care. Am J Respir Crit Care Med. 2015;192(9):1045–51.

11. Sackett DL. Superiority trials, non-inferiority trials, and prisoners of the 2-sided null hypothesis. Evid Based Med. 2004;9(2):38.

12. Prescott HC, Calfee CS, Thompson BT, Angus DC, Liu VX. Toward smarter lumping and smarter splitting: rethinking strategies for sepsis and acute respiratory distress syndrome clinical trial design. Am J Respir Crit Care Med. 2016;194(2):147–55.

13. Gordon AC, Mason AJ, Thirunavukkarasu N, Perkins GD, Cecconi M, Cepkova M, Pogson DG, Aya HD, Anjum A, Frazier GJ, et al. Effect of early vasopressin vs norepinephrine on kidney failure in patients with septic shock: the VANISH randomized clinical trial. JAMA. 2016;316(5):509–18.

14. Gordon AC, Perkins GD, Singer M, McAuley DF, Orme RM, Santhakumaran S, Mason AJ, Cross M, Al-Beidh F, Best-Lane J, et al. Levosimendan for the prevention of acute organ dysfunction in sepsis. N Engl J Med. 2016;375(17):1638–48.

15. McAuley DF, Laffey JG, O'Kane CM, Perkins GD, Mullan B, Trinder TJ, Johnston P, Hopkins PA, Johnston AJ, McDowell C, et al. Simvastatin in the acute respiratory distress syndrome. N Engl J Med. 2014;371(18):1695–703.

16. Russell JA, Walley KR, Singer J, Gordon AC, Hebert PC, Cooper DJ, Holmes CL, Mehta S, Granton JT, Storms MM, et al. Vasopressin versus norepinephrine infusion in patients with septic shock. N Engl J Med. 2008;358(9):877–87.

17. Rochwerg B, Oczkowski SJ, Siemieniuk RAC, Agoritsas T, Belley-Cote E, D'Aragon F, Duan E, English S, Gossack-Keenan K, Alghuroba M, et al. Corticosteroids in sepsis: an updated systematic review and meta-analysis. Crit Care Med. 2018;46(9):1411–20.

18. Annane D, Renault A, Brun-Buisson C, Megarbane B, Quenot JP, Siami S, Cariou A, Forceville X, Schwebel C, Martin C, et al. Hydrocortisone plus fludrocortisone for adults with septic shock. N Engl J Med. 2018;378(9):809–18.

19. Venkatesh B, Finfer S, Cohen J, Rajbhandari D, Arabi Y, Bellomo R, Billot L, Correa M, Glass P, Harward M, et al. Adjunctive glucocorticoid therapy in patients with septic shock. N Engl J Med. 2018;378(9):797–808.

20. Annane D, Sebille V, Charpentier C, Bollaert PE, Francois B, Korach JM, Capellier G, Cohen Y, Azoulay E, Troche G, et al. Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. JAMA. 2002;288(7):862–71.

21. Calfee CS, Delucchi KL, Sinha P, Matthay MA, Hackett J, Shankar-Hari M, McDowell C, Laffey JG, O'Kane CM, McAuley DF, et al. Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. Lancet Respir Med. 2018;6(9):691–8.

22. Rothenberg FG, Clay MB, Jamali H, Vandivier-Pletsch RH. Systematic review of beta blocker, aspirin, and statin in critically ill patients: importance of severity of illness and cardiac troponin. J Investig Med. 2017;65(4):747–53.

23. Knaus WA, Harrell FE Jr, LaBrecque JF, Wagner DP, Pribble JP, Draper EA, Fisher CJ Jr, Soll L. Use of predicted risk of mortality to evaluate the efficacy of anticytokine therapy in sepsis. The rhIL-1ra Phase III Sepsis Syndrome Study Group. Crit Care Med. 1996;24(1):46–56.

24. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818–29.

25. Shankar-Hari M, Harrison DA, Rowan KM, Rubenfeld GD. Estimating attributable fraction of mortality from sepsis to inform clinical trials. J Crit Care. 2018;45:33–9.

26. Ioannidis JPA, Lau J. Heterogeneity of the baseline risk within patient populations of clinical trials - a proposed evaluation algorithm. Am J Epidemiol. 1998;148(11):1117–26.

27. Burke JF, Hayward RA, Nelson JP, Kent DM. Using internally developed risk models to assess heterogeneity in treatment effects in clinical trials. Circ Cardiovasc Qual Outcomes. 2014;7(1):163–9.

28. Senn S. Disappointing dichotomies. Pharm Stat. 2003;2(4):239–40.

29. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ. 2015;350:h2147.

30. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95(Suppl 1):S144–50.

31. Ioannidis JP, Cappelleri JC, Schmid CH, Lau J. Impact of epidemic and individual heterogeneity on the population distribution of disease progression rates. An example from patient populations in trials of human immunodeficiency virus infection. Am J Epidemiol. 1996;144(11):1074–85.

32. Semler MW, Self WH, Wanderer JP, Ehrenfeld JM, Wang L, Byrne DW, Stollings JL, Kumar AB, Hughes CG, Hernandez A, et al. Balanced crystalloids versus saline in critically ill adults. N Engl J Med. 2018;378(9):829–39.

33. McKown AC, Huerta LE, Rice TW, Semler MW, Pragmatic critical care research G. Heterogeneity of treatment effect by baseline risk in a trial of balanced crystalloids versus saline. Am J Respir Crit Care Med. 2018;198(6):810–3.

34. Shankar-Hari M, McAuley DF. Divide and conquer: identifying acute respiratory distress syndrome subphenotypes. Thorax. 2017;72(10):867–9.

35. Russell JA, Lee T, Singer J, Boyd JH, Walley KR, Vasopressin SSTG. The septic shock 3.0 definition and trials: a vasopressin and septic shock trial experience. Crit Care Med. 2017;45(6):940–8.

36. Wong HR, Atkinson SJ, Cvijanovich NZ, Anas N, Allen GL, Thomas NJ, Bigham MT, Weiss SL, Fitzgerald JC, Checchia PA, et al. Combining prognostic and predictive enrichment strategies to identify children with septic shock responsive to corticosteroids. Crit Care Med. 2016;44(10):e1000–3.

37. Wong HR, Cvijanovich NZ, Anas N, Allen GL, Thomas NJ, Bigham MT, Weiss SL, Fitzgerald J, Checchia PA, Meyer K, et al. Developing a clinically feasible personalized medicine approach to pediatric septic shock. Am J Respir Crit Care Med. 2015;191(3):309–15.

38. Antcliffe DB, Burnham KL, Al-Beidh F, Santhakumaran S, Brett SJ, Hinds CJ, Ashby D, Knight JC, Gordon AC:Transcriptomic Signatures in Sepsis and a Differential Response to Steroids. From the VANISH Randomized Trial. Am J Respir Crit Care Med. 2019;199(8):980-86.

39. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, Clarke M, Gargon E, Gorst S, Harman N, et al. The COMET handbook: version 1.0. Trials. 2017;18(Suppl 3):280.