Equivalent Efficacy of Hydroxyethyl Starch 130/0.4 and Human Serum Albumin

If Nothing Is the Same, Is Everything Different? The Importance of Context in Clinical Trials and Statistics

During the past year, there has been substantial discussion about hydroxyethyl starches (HESs), with some claiming that neither efficacy nor safety has been demonstrated.

In this issue of Anesthesiology, Van der Linden et al. report a clinical efficacy trial with HES 130/0.4 that has several interesting issues. Some, but not all, are discussed relatively briefly, as space limitations preclude full exposition. Only tetrastarches (HES 130/0.4 and 130/0.42) are discussed, as older HESs with larger molecules and greater degree of substitution are outmoded owing to their undesirable effects, most notably, on coagulation and hemostasis and tissue accumulation.

First, is the importance of the main finding of Van der Linden et al.: an efficacy of HES 130/0.4 equivalent to that of 5% human serum albumin when used in pediatric cardiac surgery (more about “equivalence” below: beginning with the paragraph that starts with “In sub-titling...”). This extends the data from other trials in adults, assessing the efficacy ..., and in doing so demonstrating volume expansion similar to that of other colloids. Of some importance, as well, was the secondary finding of a lesser positive fluid balance with HES 130/0.4. Lesser positive fluid balance has produced improved outcomes; however, this trial was neither designed nor powered to assess that.

Safety data are also presented by Van der Linden et al. There were no differences between groups for mortality, biomarkers of renal injury or function, coagulation tests, transfusion of any blood component, or calculated blood loss. Despite the 28-day follow-up and no indication of safety concerns, it is not possible to draw confident conclusions regarding safety from the current publication, as with an HES safety sample size of 31 per group, the power to detect differences was small, and the absence of a specific event (e.g., death) is compatible with a 95% upper confidence limit of an incidence of 9.2%.

Data regarding safety of HES products are of considerable importance, as recently two randomized, “pragmatic” trials in critically ill patients in intensive care units purported to indicate some safety issues with two different tetrastarches (130/0.42 in the Scandinavian Starch for Severe Sepsis [“6S”] trial and 130/0.4 in the Crystalloid versus Hydroxyethyl Starch trial [“CHEST”]). However, a number of issues cloud the interpretation of the results of those trials. Trial therapy was initiated many hours after intensive care unit admission (approximately 4 h in 6S and 10 h in CHEST), after resuscitation, with fluid boluses with no objective criteria for either initiation or termination of test fluid administration. Central venous pressure at initiation of test fluid therapy was approximately 10 mmHg in both trials and greater after HES administration. The 6S trial enrolled...
patients with severe sepsis with approximately 30% in shock, and the CHEST trial had a substantial fraction of similar patients. Fluid therapy was not goal directed and did not take into consideration the pharmacokinetic differences of the test fluids14,15 administered.

An intact endovascular glycocalyx functions as a selective filter and is critical for the intravascular retention of fluids and solutes, preventing their extravasation into extravascular compartments.16–18 Degradation of the glycocalyx results in immediate tissue edema.19 Microscopic observation in laboratory experiments and measurements of function and concentrations of glycocalyx degradation compounds in plasma in laboratory experiments and humans, including patients in intensive care units, indicate that the glycocalyx is degraded by sepsis, septic shock, endotoxemia, hypoxia, ischemia, and increased atrial natriuretic peptide secretion.18–30 Sepsis and shock likely had degraded the endovascular glycocalyx in the CHEST and 6S trials, with further degradation induced by atrial natriuretic peptide secretion27,30 owing to additional atrial stretching secondary to colloid administration to an already full atrium. A degraded glycocalyx diminishes or eliminates the efficacy of an administered colloid by failing to retain it within vessels. The resulting presence of a colloid in the extravascular space might result in unintended consequences. In the presence of an intact glycocalyx, there is far less tissue accumulation of modern tetrastarches (as tested by Van der Linden et al.) than older HES compounds of higher molecular weight and greater molecular substitution.31

Both CHEST and 6S trials had mortality as the primary endpoint. The former found no difference in mortality with HES 130/0.4, and the latter found increased 90-day mortality with HES 130/0.42 (P = 0.040*) in the intention-to-treat population. However, the statistical difference did not endure in the two per-protocol analyses (those patients actually treated with the test fluids; P = 0.07 and P = 0.12) or analyses of the Kaplan–Meier survival curves (P = 0.07 and P = 0.14). I have always thought it more logical to analyze randomized, blinded trials for efficacy according to the articles actually administered rather than according to theoretical intent.

CHEST and 6S trials reported increased need for renal replacement therapy (P = 0.0495 and P = 0.047, respectively) although there were no a priori criteria for institution of this therapy, and it was not statistically significant in CHEST after correction for covariates. Moreover, it was not supported by adverse differences of Risk of renal dysfunction; Injury to the kidney; Failure of kidney function; Loss of kidney function; and End-stage kidney disease (RIFLE) criteria in either trial, which is the accepted standard for assessment of renal impairment32,33 or renal replacement therapy + sepsis-related organ failure assessment score of 3 or greater, or increased creatinine two times or more from baseline in 6S. Neither trial corrected these findings for fluid balance, which can result in a falsely positive finding of acute kidney injury in the presence of greater positive fluid balance,34 as occurred after study day 1 with HES in CHEST. Perhaps, the further increase in central venous pressure when HES was given to an already full atrium, with likely greater intravascular retention than with crystalloid, contributed to the subjective decision of instituting renal replacement therapy despite lack of RIFLE criteria to support that decision. The striking difference of the patient populations and the purpose and timing of therapy of those trials bear little relationship to the use of HES in the operating room, as originally intended, for blood volume restoration. There is no evidence to support the use of any colloid for fluid bolus in septic patients; rather there is suggestion of harm.35 Similarly, the CHEST and 6S trials might hint at some safety issues with tetrastarches when used in that manner in the intensive care unit. One should be cautious in drawing conclusions regarding the use of pharmaceuticals for a context differing from that tested.

With respect to intraoperative use, two recent reviews of published trials of tetrastarches used to treat a volume deficit in the perioperative period36,37 have concluded a lack of tetrastarch-induced mortality,36 adverse effects on renal function,36,37 or clinical hemostasis,36 while noting a relatively brief duration of clinical follow-up for most trials.† Van der Linden et al.1 now add data for a substantial postoperative period, as seen with some,38,39 but not many reports, in comparing the efficacy of HES 130/0.4 with that of other fluids. The primary efficacy endpoint was that of colloid volumes used during pediatric cardiac surgery. HES and human serum albumin were used in “equivalent” volumes, with no differences in safety measures in the approximately 30 patients per group, through 28 postoperative days. This efficacy “equivalence” was established firmly and is in keeping with pharmacokinetic knowledge in humans40–43 and understanding of the function of an intact endovascular glycocalyx.16,17,19 Data from perfused ischemic rat and guinea pig hearts and humans undergoing cardiopulmonary bypass suggest glycocalyx degradation during these conditions. However, it is difficult to know whether that occurred in the children studied by Van der Linden et al., as the absence of a crystalloid control group prevents concluding the absence of any effect despite the lack of difference of adverse events and serious adverse events between the tetrastarch and human serum albumin.
Considering the extant data and their limitations, it is probably prudent to avoid the use of HES for intermittent boluses after resuscitation in patients with sepsis, and especially septic shock. Similarly, it might be sensible to avoid the use of HES for fluid maintenance for other conditions of ischemia/hypoxia. A notable exception is that of trauma, where a full study and a subgroup analysis of CHEST did not demonstrate a detrimental effect of a tetrastarch. There is a lack of data to cause concern for the use of tetrastarches for perioperative volume replacement. Additional data should be available when the recently completed and presented study testing Colloids versus Crystalloids for Resuscitation in the Critically Ill trial (“CRIStAL”) is published.

Those involved in this phase 4 trial are to be commended. Regulatory authorities impose such a requirement when it is considered that efficacy and safety data are sufficient for approval, but that additional information is desirable or necessary for a more complete understanding of the risk-benefit profile. However, a substantial majority of phase 4 trials are never initiated, fewer are completed, and even less are reported in the public domain.

In subtitling this editorial, “If nothing is the same, is everything different?” I have paraphrased Hanley and Lippman-Hand’s classic article “If nothing goes wrong, is everything all right?”11 Here, I point out, very briefly, hopefully for the benefit of the readership, conceptual, and practical differences between trials of “superiority,” “noninferiority,” and “equivalence.” Demonstration of superiority of one therapy compared with another requires rejection of the null hypothesis that the groups do not differ statistically. That is, the probability that there is no difference is less than an a priori arbitrarily defined value, usually, by convention, 5% (\(P < 0.05\)). In such a test, there is no a priori (or at times, even post hoc) definition of what constitutes a clinically important difference. Thus, statistical significance can result in a difference without clinical meaning (“a tale . . . full of sound and fury, signifying nothing.”45§) (for example, the CHEST trial reported difference \([P < 0.001]\) of transfusion of unspecified “blood products” of 0.057 ml kg\(^{-1}\) day\(^{-1}\) for the first 4 days). Therein lies the conceptual difference between on the one hand a “superiority” trial and on the other hand “noninferiority” and “equivalence” trials. In the latter two, the quantitative difference between groups that determines trial success or failure must be decided a priori. This difference, frequently labeled as “delta” (\(\Delta\)), or sometimes “M,” should have a clinically meaningful basis. The magnitude of this limit is subject to opinion or may be imposed by regulatory authorities, as occurred in the trial reported by Van der Linden et al.1 Furthermore, in these trials, the test article should be assessed against a comparator with proven value; without that, there is little clinical meaning to “noninferior” or “equivalent.”11 The success or failure of these trials rests not only on the mean difference between groups but also on the CI (by convention, usually 95%) associated with that mean. The \(\Delta\) should be chosen with great care, as when a limit of the CI lies outside \(\Delta\) by even a small amount, the result is a failed trial18 that can lead to absence of regulatory approval and termination of product development. Figure 1 depicts possible results of noninferiority testing. To demonstrate noninferiority, the lower confidence limit must be above the a priori designated −\(\Delta\). Examples 1A, B, D, and E show non-inferiority, whereas 1C does not, because although the mean difference is above the specified −\(\Delta\), it is not so with 95% confidence or greater, as the lower 95% CI limit is below −\(\Delta\). Noninferiority design, scientific, statistical, and regulatory considerations are reviewed in depth elsewhere.49#

A clinical trial test of “equivalence” is somewhat analogous with respect to the imposition of an a priori delineated \(\Delta\) but is a bidirectional test. That is, the mean difference between groups and both sides of the CI must be within ±\(\Delta\). In figure 2, examples 2A and B would be declared to be “equivalent” (the Van der Linden trial fits these), but 2C, D, and E would not, as the 95% CI band of each extends outside ±\(\Delta\). Note that 2C would also fail tests of noninferiority and superiority. However, 2D and E are clearly noninferior, but are not “equivalent.” 2D has the possibility of having a difference greater than ±\(\Delta\) and thus includes the possibility of superiority. 2E actually demonstrates superiority, as the
mean and the entire 95% CI exceed zero. It should be noted that regulatory authorities recognize the ability to switch, post hoc, between analyses for noninferiority and superiority, provided the trial design is adequate, the primary outcome variable, and the Δ remain unchanged, and the possibility is included in the a priori statistical analysis plan.50 “Equivalence” is not ordinarily included in this switching paradigm; however, many trials designated as “equivalence” trials are really tests of “noninferiority”# (but not in the case of the trial reported by Van der Linden et al.1). “Superiority” of efficacy might also suggest the possibility of differences of safety, owing to greater potency or a different mechanism of action. Thus, a comparison could fail to be “equivalent,” as in 2D, because the test article might (P > 0.05) actually be better, creating an indeterminate state. It is possible to be unable to accept the hypothesis that there is equivalence (that there is no statistical difference: the 95% CI of each extend outside ±Δ), and F fail the test of equivalence, as the 95% CI of each extend outside ±Δ. F presents a possible trial result appearing to demonstrate simultaneously both equivalence and superiority, highlighting the importance of context.

**Fig. 2.** Possible results of tests of equivalence. −Δ and +Δ are the a priori assigned equivalence margins. The results of Van der Linden et al.1 fit examples A or B. Examples A and B pass the test of equivalence, as the mean and 95% CI of each are within ±Δ. Examples C, D, and E fail the test of equivalence, as the 95% CI of each extend outside ±Δ. F presents a possible trial result appearing to demonstrate simultaneously both equivalence and superiority, highlighting the importance of context.

a priori definition of the quantitative difference that constitutes clinical importance. However, in an equivalence trial, in the light of a clinically determined a priori Δ, that probable difference would be assessed to be of no clinical importance, and hence “equivalence” will have been established (provided other conditions are met): the difference, although established by rejection of the null hypothesis of a lack of difference with statistical probability, is so small as to be not clinically meaningful (i.e., is “equivalent”).

It is interesting to see a report of an equivalence trial for a mature product, such as HES 130/0.4. An increasing number of equivalence and noninferiority trials are to be expected, as governmental regulations and rules are slowing pharmaceutical development, and a growing number of generic, biosimilar, and “me-too” products are likely.

Richard B. Weiskopf, M.D., University of California, San Francisco, California. rbw@itsa.ucsf.edu, rbw@theweiskopfgroup.com

References


37. Martin C, Jacob M, Vicaut E, Guidet B, Van Aken H, Kurz A: Effect of waxy maize-derived hydroxyethyl starch 130/0.4 and HES 200/0.5 after repeated intravenous administration in rats. Drugs R 2003; 4:331–8


