Must hypervolaemia be avoided? A critique of the evidence

Robert G. Hahn

Research Unit, Södertälje Hospital, Södertälje, Sweden

Abstract
Anaesthetists are cautioned to avoid hypervolaemia in their patients. The most cited reason is that hypervolaemia elicits the release of atrial natriuretic peptides that damage the endothelial glycocalyx layer. Although shedding of the glycocalyx causes extravasation of protein in inflammatory disorders, it is more uncertain whether hypervolaemia alone is enough to cause clinically important shedding.

This review scrutinises the methodology used in two key papers that propose such a link. The most cited one reports that hydroxyethyl starch and 5% albumin, when creating a hypervolaemic state, only expands the plasma by 40% of the infused volume. This result was obtained by comparing measurements of the plasma volume performed with the indocyanine green (ICG) dye method before and after the infusion. However, the transit time of the dye, as well as inequality in the concentration between vascular beds, both act to underestimate the plasma volume, particularly as times were extrapolated backwards to time zero instead of to (the more correct) 1 minute.

A recalculations based on theoretical ICG data, taking account of the transit time, shows the plasma volume expansion was closer to 100% than to 40% of the infused volume. This figure is supported by the dilution of the reported blood haemoglobin and plasma protein concentrations, as well as by other sources.

In conclusion, only weak evidence supports a fluid-induced release of atrial peptides of sufficient size to alter the kinetics of colloid fluid by shedding of the endothelial glycocalyx layer.

Key words: clinical pharmacology; fluid therapy, therapeutic use; glycocalyx, physiology; hydroxyethyl starch, pharmacokinetics; indicator dilution technique, method

Several aspects of the current recommendations for fluid therapy are confusing. The anaesthetist is told to use a goal-directed protocol, which usually means the creation of hypervolaemia by the use of fluids to increase the stroke volume. However, virtually all evaluations of goal-directed fluid therapy are based on titration with 6% hydroxyethyl starch 130/0.4 (HES), which authorities advise us not to use. Hypervolaemia should also be avoided because it raises the plasma concentrations of atrial natriuretic peptides, thereby promoting protein extravasation by damaging the vascular endothelium [1, 2]. Thus, the anaesthetist may have difficulty integrating these conflicting suggestions. This review article will discuss the evidence behind the warnings about hypervolaemia.

HYPERVOLAEMIA DURING SURGERY
The first issue to address is whether hypervolaemia is needed during anaesthesia, surgery and intensive care. My own answer is yes. Many diseases cause vasodilatation, which shifts blood from the “stressed” to the “unstressed” blood volume. Vasodilatation is also the most prominent vascular effect of regional and general anaesthesia.

Infusion fluids are the ultimate first-line treatment for the arterial hypotension and impairment of organ perfusion arising as consequences of this vasodilatation. A flow-guided volume optimisation protocol typically places the patient on a hypervolaemia of about 1 L, where the variability is mainly determined by the degree of arterial hypotension [3]. If anaesthesia is given without fluid, the plasma volume will rise slowly by means of a spontaneous capillary refill process [4]. Hypervolaemia is necessary to maintain organ perfusion in this setting, regardless of whether it is created by infusion fluids or by capillary refill.

Although much of the infused fluid might be confined to the “unstressed” blood volume, a rise in the central venous pressure is still the expected consequence of flow-guided
volume optimisation that aims to increase stroke volume in the anaesthetized patient [3].

**DANGERS OF HYPAVOLAEIA**

Hypervolaemia has a number of well-known physiological effects in our patients. Distribution of crystalloid fluid from the plasma to the interstitial fluid space occurs rapidly, as the rate is proportional to the degree of blood volume expansion [5]. Hence, peripheral oedema develops faster than it does when fluid is infused slowly. Hypervolaemia also increases myocardial work and cardiac pressures, at least when the degree of anaesthesia-induced vasodilatation is exceeded. Moreover, crystalloid fluid decreases the colloid osmotic pressure that, together with raised pressures, promotes pulmonary oedema [6].

In recent years, the warnings of hypervolaemia have focused on another sequence of events; namely, that it causes release of atrial natriuretic peptides (ANPs) that break down the endothelial glycocalyx layer [1, 2]. Shedding of the glycocalyx promotes protein and fluid leakage from the plasma. This reduces the effectiveness of colloid fluids and might explain, at least in part, the peripheral oedema that accompanies surgery and intensive care. The ANP-glycocalyx link has also been used to account for the perceived poor effectiveness of HES in some clinical studies [7, 8].

**ATRIAL NATRIURETIC PEPTIDES**

Raised plasma concentrations of ANP in response to hypervolaemia were first described in the early 1980s. These peptides are released from the right atrium and they reduce the blood volume by promoting natriuresis and the extravasation of albumin [9, 10].

In 1990, Kamp-Jensen et al. showed that the ANP concentration doubles in response to a brisk intravenous infusion of 2 L of crystalloid fluid [11]; this finding has later been corroborated by others [12, 13].

Plasma ANP does not seem to increase after infusion of smaller volumes of crystalloid fluid. My own group administered 5 and 10 ml/kg Ringer’s acetate over 15 min in healthy volunteers, and found no rise in plasma ANP in any of them. Although these measurements were performed as part of a larger study of fluid kinetics, they were not reported there [14].

Stable precursors of ANP have become clinical tools in the diagnosis and treatment of heart failure [15]. Patients with heart failure show a much greater rise than is seen with infusion fluids; the increase is typically 50-fold. Only a part of these differences can be attributed to the fact that the precursors have a longer half-life.

**HYLAURONAN AND SYNDECAN-1**

As early as 1994, Berg et al. showed that the plasma hyaluronan (hyaluronic acid) concentration doubles in response to a 45-min infusion of as little as 1 L of crystalloid fluid [16]. This rise was believed to indicate a wash-out of hyaluronan from the interstitial fluid space, where it is a main constituent.

Later researchers have interpreted a rise in plasma hyaluronan, together with syndecan-1 and heparan sulphate, as evidence of the release of constituents from the luminal surface of the endothelium. Although the glycocalyx layer is damaged from inflammation and high ANP levels [17], it is less clear whether the doubled ANP level that develops in response to fluid infusions is sufficient to create an effect that warrants generally held warnings of hypervolaemia.

Chappell et al. made this claim by reporting an 80% rise in plasma ANP, hyaluronan and syndecan-1 concentrations after an infusion of 1 L of HES in anaesthetised patients awaiting surgery [2]. However, the measured plasma concentrations probably changed very little, if at all, because the reported concentrations were corrected for plasma albumin, which becomes strongly diluted by (in particular) the first fluid infused after general anaesthesia has been induced. A study by my research group confirmed that the haemodilution arising from HES in that setting corresponded to expansion of the plasma volume by twice as much as the infused volume (500–600 mL) [3]. The same albumin dilution (40–50%) would cancel out all changes in plasma concentrations reported by Chappell et al. [2].

Biological effects are, with few exceptions, linked to the measured plasma concentration. Adjusting the plasma concentrations for albumin actually converts the data to an index of their total amounts in the circulation. Without further exploration, the implications of such data are unclear, as an increase in plasma content of a biomarker can be due increased production, decreased elimination or simply equilibration with the extravascular concentration.

**POOR INTRAVASCULAR PERSISTENCE OF HES**

The key publication used to support the concept of a detrimental effect of hypervolaemia on the efficacy of colloid fluid also stems from Chapell’s group, and was published in 2001. Rehm et al. measured the plasma volume by indocyanine green (ICG) before and after infusion of HES in anaesthetised patients awaiting surgery [18]. Only 40% of an infusion of 1.4 L of HES was found responsible for expanding the plasma volume. The remainder of the infused fluid was thought to have immediately left the bloodstream, presumably due to a hypervolaemia-associated effect on the ANP-glycocalyx axis. Thus far, I have counted five re-
-publications of their main illustration showing the 40% efficacy of the infused HES volume, compared to 100% when hypervolaemia is not induced [19–23]. The list even includes a prominent Clinical Update in *The Lancet* [19]. At the 2014 IFAD meeting, I understood that the adverse effect of the ANP-glycocalyx link on the efficacy of colloid fluid has gained wide acceptance among experts in the field. Their argument was that the results had been obtained with the best possible methodology [22]. For that reason, I felt that a critical review of the evidence is warranted.

**THE INDOCYANINE METHOD**

The study by Rehm *et al.* reported plasma volumes measured with indocyanine green (ICG), a dye that binds to albumin [18]. A bolus was injected into a central vein and the arterial plasma concentration measured 2, 3, 4 and 5 min later. The plasma volume was obtained as the dose divided by the concentration that had been extrapolated back to zero time from the mono-exponential arterial time-concentration curve. A very large volume (1.3–1.4 L) of hetastarch (MW 200) or 5% albumin was then infused over 15 min and the plasma volume measured again 30 min later.

Although this research methodology may seem rational, it is not. The reason is that ICG is eliminated almost exclusively in the liver, and when the calculations indicate the fastest elimination (i.e. during the first minute), the ICG has not even reached the liver. The slope of the ICG curve is mirrored at an erroneously high ICG concentration at time zero because elimination, in reality, begins at 1 minute (Fig. 1). Hence, overlooking the transit time for the dye from the site of injection to the site of elimination makes all plasma volumes too small.

**IS POOR VOLUME EXPANSION AN ARTEFACT?**

Polidori and Rowley [24] have shown that the plasma volume is underestimated by 25% in a typical patient if extrapolation is done to zero time instead of the (more correct) 1 min. The error is negligible if the ICG clearance is very low, but gradually becomes larger as the clearance increases. Polidori and Rowley state that the reasons include both a steeper elimination slope for the concentration-time curve and greater differences in dye concentration between the hepatic and extra-hepatic circulatory loops.

The ICG clearance is a scientifically accepted method for measuring the liver blood flow. This flow varies with cardiac output, which is reduced by about 35% when general anaesthesia is induced without any accompanying fluid [3]. The ICG clearance is likely to be low in such low-flow states. However, the transit time for the dye will also be prolonged if the ICG clearance falls due to a reduction of cardiac output. Therefore, any underestimation of the plasma volume in a patient will persist as a constant proportional error as long as liver blood flow changes in proportion to the transit time. Nevertheless, a stable proportion of 1:1 between these entities can be difficult to guarantee when inducing dramatic changes in haemodynamics, which is the case when inducing general anaesthesia followed by infusion of a massive amount of colloid fluid over 15 min.

**HOW THE ERROR DEVELOPS**

Polidori & Rowley’s article shows that we must be open to the possibility that the poor volume effect of colloid fluids in hypervolaemia is the result of a miscalculation. We can study how such an error may have developed in the following theoretical example.

Assume that the half-life of the ICG is 3 min. The ICG bolus reaches the liver after 1 min and its initial concentration is reduced to 50% at 4 min and to 25% at 7 min. Taking the natural logarithm of these fictuous concentration data yields 4.836 at zero min and 4.605 at 1 min. Back-transformation to the linear scale shows that the ICG concentration at time zero is 79% of the (correct) one at 1 min. The true plasma volume should then be 26.6% higher than the volume indicated by extrapolation to zero time.

Although the baseline plasma volume obtained the ICG by Rehm *et al.* was 2.98 L, it should, if extrapolation had been made to 1 min, have been 3.77 L. Assume that we infuse 1.4 L of HES which entirely remains in the plasma, except for consideration of the normal half-life of 2 hours [25]. Thirty minutes later the true plasma volume would then be 3.77 + (1.4*0.85) = 4.96 L. Based on this volume, the error of 26.6% now represents 1.32 L, i.e. 530 mL more than before the infusion. Back-extrapolation of new ICG curve to zero time,
as was done by Rehm et al., would now show 4.96−1.32 = 3.64 L, which is virtually identical to the volume reported in their original publication.

To sum up, the new calculations are based, firstly, on the idea that HES expands the plasma volume by 100% of the infused volume; secondly, that the rate of elimination of the HES volume is perfectly normal as given by other works; and, thirdly, that account is taken of the transit time for the dye between the site of injection and the liver.

The obtained result does not give reason to suggest that the volume effect of colloid fluid is poor when given in the hypervolaemic state.

**PROTEIN AND HAEMATOCRIT DILUTION**

The original paper by Rehm et al. [18] contains other data suggesting that the volume effect of HES in hypervolaemic volunteers was indeed much higher than 40%. The plasma protein concentration decreased from 64 to 44 g L⁻¹, which corresponds to an increase in the plasma volume by (64−44)/44 = 45%. By assuming a plasma volume at a baseline of 3 L, we understand that a plasma volume expansion of 45% allows all the infused 1.3 L of HES to remain in the bloodstream. This calculation assumes that only very minimal protein leakage occurred, which seems to be the case in this setting [26].

The paper also reports data on the haematocrit, which is a useful comparator because red blood cells are too large to leak extravascularly. The plasma volume expansion based on the reported mean haematocrit values was (35.3−27.5)/27.5/(100−35.3) which is 45%; i.e. identical to the protein dilution. This supports the intravascular retention of the entire infused HES volume. The close similarity between the dilution of the haematocrit and the protein concentration also shows that capillary leakage of macromolecules must have been very small in this setting, which is inconsistent with the presence of a marked ANP-induced increase of the capillary permeability.

Another problem is an ex vivo study that showed the preservation of endothelial integrity by 5% albumin [27]. However, the intravascular retention of the albumin was also found to be as poor as that described for HES when measured by Rehm et al. with the ICG method [18]. This suggests that one of the methodologies used is incorrect.

Hence, there are reasons to believe that the reported 40% volume efficacy could as well be close to 100%. We may conclude that ICG is a soft method for the determination of the plasma volume in situations where the plasma volume has dramatically changed. A key problem is that the error inflicted by extrapolating to zero time is proportional, while reported changes in plasma volume are absolute. Methodological uncertainties arise when the cardiac output is changed. Although the results would be more trustworthy if extrapolation backwards is made to 1 minute instead of to zero time, comparisons of two measurements can be done safely only when cardiac output has been left unchanged.

Scepticism to the accuracy of the ICG method has previously been expressed following the use of a physical mixing apparatus [28]. Jacob et al. have also studied the accuracy of the ICG method, but then focused on what happens when the sampling time is extended to beyond 5 min, rather than on the issues considered here [29].

**THE VOLUME KINETIC METHOD**

Although in the 1980s [30], as well as later [31], I personally worked with tracers to estimate body fluid volumes, I try to refrain from using them as they are subject to a host of potential errors, some of which may not even be realised today. As an alternative, I have developed volume kinetics as a robust alternative method for studying the behaviour of infusion fluids in the living human body.

This approach is based on 20–35 measurements of the haemodilution resulting from an infusion fluid. The transit time for the fluid is not important, since the measurements are extended over 3–5 hours. The only assumption, beyond those made in the kinetic model used for analysis, is that the haemoglobin (Hb) molecules are evenly distributed, on average, in an expandable body fluid compartment during that period of time.

Kinetic calculations of fluids based on Hb show the distribution of the fluid volumes and not the blood volume. The size of the Hb pool is even without relevance. However, results can be misleading results if the Hb mass is changed during the course of an experiment, which is the case in haemorrhage. Corrections for such alterations may be applied, if known.

Figure 2A shows an example of how the effectiveness of HES for expanding the plasma volume can be calculated based on careful measurements of the haemodilution. Extrapolation of the curve back to time zero yields a plasma dilution of 27% from an infusion of HES. The volume of distribution is then calculated as the infused volume (0.8 L) divided by 27%. This ratio yields 3 L, which shows that HES expands the plasma volume by 100%, as 3 L was the expected size of the plasma volume at the baseline in these volunteers. Estimates with greater precision can be obtained by fitting mathematical equations to the data [5].

**VOLUME KINETICS OF COLLOID FLUIDS**

Three volume kinetic studies have been published on conscious volunteers made markedly hypervolaemic with colloid fluid. The cardiac pressures in these volunteers were not reduced by general anaesthesia, as they were in the studies by Chappell et al. [2] and Rehm et al. [18], a factor
which more clearly challenges the influence of fluid volume on the ANP-glycocalyx axis.

The studies comprised 6% dextran 70 [32], 5% albumin [33], and HES [25]. The pooled data from all 32 volunteers has now been recalculated by using modern kinetic software for nonlinear mixed effects (Phoenix NLME, Pharsight, St. Louis, MO). The kinetic constants obtained for these colloids are illustrated in Figures 2B and 2C.

Subplot B shows that dextran 70 initially attracts some fluid from the interstitial fluid space, an aspect which is also addressed by the manufacturer. In contrast, 5% albumin has a volume of distribution that is 20–25% larger than the expected size of the plasma volume, which might be due to the ease by which albumin enters the sinusoids of the liver. HES has a volume of distribution that is almost identical to the plasma volume.

The constant shown in Fig. 2C represents the slope of the elimination curve. The half-life of the plasma volume expansion is the natural logarithm of 2 (0.693) divided by this elimination rate constant. This appears to correlate closely with the urinary excretion.

### FLUID BOUND TO THE GLYCOCALYX LAYER

Several attempts have been made to invalidate haemodilution as an approach to assess the behaviour of infusion fluids in the body. The most recent attempt claims that haemodilution is erroneous because the method does not take into account the changes in the fluid volume bound by the glycocalyx along the endothelial lining [17, 22].

This proposal rests heavily on the same study by Rehm et al. [18], who reported that 700 mL of protein-free fluid resided in the endothelial surface layer (ESL) after induction of general anaesthesia. Two thirds of it is then washed out to the hemoglobin-mixed plasma in response to the infusion of HES. These authors’ view is that the plasma volume increases by 40% of the infused HES volume, while the circulating plasma volume still increases by twice as much, by virtue of this massive washout of fluid residing in the glycocalyx layer.

These calculations seem to have been widely adopted by the medical community. However, the magnitude of the washout is directly dependent on the same erroneous calculation of the post-infusion plasma volume with ICG as discussed previously. The 700 mL derived from these ICG measurements have even been further extrapolated to indicate that the glycocalyx layer has an average thickness of 2 µm in the human cardiovascular system [17].

### THE HAEMATOCRIT FACTOR

A review of the equations used by Rehm et al. [18] to demonstrate the large amount of plasma proposed to be hidden in the glycocalyx layer is the volume corresponding to the haematocrit factor, which is an expression of the difference in haematocrit between whole-body and peripheral blood. The background is that a blood volume measured with a double-isotope method (i.e. where the erythrocyte and plasma volumes are measured by different tracers) has to be multiplied by 0.91 to indicate the same average erythrocyte/blood volume ratio as the directly measured haematocrit.

As early as 1953, the haematocrit factor in humans was calculated as 0.910, with a standard deviation of 0.024 [34]. Rehm et al. [18] overlooked performing the conventional corrections for both the haematocrit factor and plasma trapping when their peripheral haematocrits were determined, and they then interpreted the summarised effect to indicate hidden fluid in the endothelial lining. Although other scientists have been cautious to give the haematocrit factor a physiological significance, many have assumed that the haematocrit really differs significantly between vascular beds. While the haematocrit has long been known to be lower in capillaries and small vessels, the blood volume...
residing there has been considered too small (< 10% of total) to account for the discrepancy [35].

A likely explanation for the haematocrit factor is that albumin, in contrast to Hb-based tracers, easily enters the liver sinusoids and thereby indicates too large a plasma volume. This view agrees well with the observation that 5% albumin has a 20–25% larger volume of distribution than the expected size of the plasma volume (Fig. 2B) and that the albumin molecule requires as much as 8 min to equilibrate completely in the plasma [36]. The haematocrit factor is also quite stable under various circumstances, such as extreme changes in haematocrit [34] and haemorrhage of between 10% and 20% of the blood volume [37], which is hardly the case for the easily-damaged glyocalyx layer. Hence, several alternative explanations can account for the observation that isotope measurements do not fully agree with the measured peripheral haematocrit.

Caution is recommended when making conclusions about the magnitude and clinical importance of fluid volumes released from the glyocalyx meshwork. Reasons include the already mentioned questionable methods used to calculate them. A frequent assumption is that all vascular beds are opened to allow blood/endothelial interaction, which is never the case. Estimates are also based on the belief that the entire glyocalyx layer is shed regardless of vessel diameter and body region. This might be true in severe disease, such as sepsis, but appears to be unlikely in hypervolaemia. Finally, plasma proteins and even sodium are partially excluded from the ESL fluid so, if released to the plasma, much of the volume will more or less immediately equilibrate across the extracellular fluid space, thereby contributing very little to the plasma volume expansion.

**KINETIC EFFECTS OF GLYOCALYX DAMAGE**

Although no valid quantification has been presented to reveal how shedding of the glyocalyx layer affects the kinetics of colloid fluid in the body, technical tools and experiences from drug kinetics can be used to outline the expected consequences.

Partial derivatives obtained during a kinetic analysis show which period of time during an experiment contributes most to the final value of each kinetic parameter. Such data suggest that release of non-circulating ESL fluid to the plasma occurring after maximum haemodilution would slightly prolong the half-life (intravascular persistence) of an infused colloid fluid volume. Firstly, this is hardly misleading if an infusion recruits fluid from a non-circulating to a circulating part of the blood volume. Secondly, recruitment of non-circulating fluid will prolong the plasma half-life of the fluid only when the calculation is based on the haemodilution, while the half-life based on urinary excretion would become shorter. As already mentioned, these two half-lives have agreed closely in the volume kinetic studies of colloid fluids where both half-lives have been possible to calculate [25, 33].

In contrast, an ANP-stimulated increase of the capillary leakage rate will reduce the half-life of a colloid fluid volume. This change will also be captured by the volume kinetic method.

None of these two alterations (release of ESL fluid/increased capillary leakage) is likely to cause a marked change in the immediate volume of distribution, which is still determined primarily by the oncotic strength of the fluid. Although colloid molecules do pass the capillary membrane more easily when the glyocalyx is damaged, one cannot expect that the fluid behaves similar to a crystalloid where this membrane is passed without any problems at all.

If a colloid fluid is infused in a subject who already has a completely shed glyocalyx layer, the expected effect on the fluid kinetics would consist of a slightly larger volume of distribution coupled with somewhat smaller haemodilution, as well as a shorter intravascular persistence time.

Although most evidence is indirect, little doubt exists that medical disease where the glyocalyx is markedly damaged have a faster turnover of colloid fluid than is seen in healthy human beings. For example, the intravascular persistence of the volume of 5% albumin closely follows the capillary leakage of albumin [33] while sepsis increases the capillary leakage rate of albumin by 300%, which would then be from 5% (normal) to 20% per hour [38]. Therefore, one can assume that sepsis reduces the half-life of the plasma volume expansion of 5% albumin from 2 hours to < 1 hour.

**FUTURE VIEWS**

The overarching question in this review is whether the clinician should be warned of hypervolaemia just because HES and 5% albumin have been shown to expand the plasma volume poorly in the presence of hypervolaemia. The evidence behind these warnings is weak and stems from being overly trusting of the calculations based on difficult tracer techniques; alternative calculations do not support the findings.

The kinetics of infusion fluids in patients with glyocalyx damage is still of great research interest in many settings, such as inflammatory disorders and postoperative care. Interestingly, the kinetics of crystalloid fluids seems to be affected by albuminuria, which indicates damage to the glyocalyces of the glomeruli. For example, minimal albuminuria increased the rate of elimination of crystalloid fluid infused just before major abdominal surgery [39] and pre-eclamptic women had twice the distribution and elimination rates of Ringer’s acetate when compared to matched pregnant controls [40].
CONCLUSIONS

1. Hypervolaemia is needed to preserve organ perfusion during regional and general anaesthesia.
2. Hypervolaemia increases the rate of distribution of crystalloid fluid, which allows a faster build-up of peripheral tissue oedema.
3. Calculations performed with the indocyanine green (ICG) method have shown that only 40% of an infused volume of hydroxyethyl starch (HES) and 5% albumin remains in the plasma volume if the patient is hypervolaemic.
4. Underestimation of the plasma volume expansion occurs if the transit time for the site of injection of ICG to the site of elimination is not taken into account.
5. Without these errors, the expected expansion of plasma volume with HES is probably close to 100%, even in hypervolaemic states; this expectation is supported by volume kinetic calculations.
6. No trustworthy information to date shows that colloid fluids would have a poor plasma volume-expanding effect just because they cause hypervolaemia in healthy human beings.

ACKNOWLEDGEMENTS

1. The authors declare no financial disclosure.
2. The authors declare do conflict of interest.

References:


Corresponding author:
Robert G. Hahn, MD, PhD
Professor of Anaesthesiology & Intensive Care
Research Unit, Södertälje Hospital
152 86 Södertälje, Sweden
e-mail: r.hahn@telia.com
Robert.hahn@sodertaljesjukhus.se