Review

Blood alcohol concentration in the clinical laboratory: a narrative review of the preanalytical phase in diagnostic and forensic testing

Cristiano Ialongo*

Department of Experimental Medicine, Policlinico Umberto I, 'Sapienza' University, Rome, Italy

*Corresponding Author: cristiano.ialongo@uniroma1.it

Abstract

The analysis of blood alcohol concentration (BAC), a pivotal toxicological test, concerns acute alcohol intoxication (AAI) and driving under the influence (DUI). As such, BAC presents an organizational challenge for clinical laboratories, with unique complexities due to the need for forensic defensibility as part of the diagnostic process. Unfortunately, a significant number of scientific investigations dealing with the subject present discrepancies that make it difficult to identify optimal practices in sample collection, transportation, handling, and preparation. This review provides a systematic analysis of the preanalytical phase of BAC that aims to identify and explain the chemical, physiological, and pharmacological mechanisms underlying controllable operational factors. Nevertheless, it seeks evidence for the necessity to separate preanalytical processes for diagnostic and forensic BAC testing. In this regard, the main finding of this review is that no literature evidence supports the necessity to differentiate preanalytical procedures for AAI and DUI, except for the traceability throughout the chain of custody. In fact, adhering to correct preanalytical procedures provided by official bodies such as European federation of clinical chemistry and laboratory medicine for routine phlebotomy ensures both diagnostic accuracy and forensic defensibility of BAC. This is shown to depend on the capability of modern pre-evacuated sterile collection tubes to control major factors influencing BAC, namely non-enzymatic oxidation and microbial contamination. While certain restrictions become obsolete with such devices, as the use of sodium fluoride (NaF) for specific preservation of forensic BAC, this review reinforces the recommendation to use non-alcoholic disinfectants as a means to achieve "error-proof" procedures in challenging operational environments like the emergency department.

Keywords: preanalytical phase; blood alcohol concentration; specimen handling; humans; substance abuse detection

Submitted: July 18, 2023 *Accepted: August 20, 2023*

Introduction

Ethyl alcohol (C_2H_5OH, CAS 64-17-5) or ethanol (EtOH) or just "alcohol", is a liquid organic compound at room temperature that forms the basis of various aqueous beverages (spirits, wine, or beer) consumed by half of the world's population over the age of 15, the majority of whom reside in the Americas, Europe, and the Western Pacific (1). For individuals up to the age of 49, the consumption of EtOH is the leading risk factor for premature death and disability, with mortality rates even exceeding those of diabetes, tuberculosis, and human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) (2).

Blood alcohol concentration (BAC) analysis is a toxicological test associated with two important conditions: acute alcohol intoxication (AAI) and driving under the influence (DUI). Acute alcohol intoxication is a clinical emergency that affects 7.5% of drinkers and causes 2200 deaths *per* year in countries such as the United States (1,3). The diagnostic determination of BAC is routinely performed by means of enzymatic reaction quantified *via* automated spectrophotometry (*i.e.*, clinical chemistry auto-analysers), using serum or plasma as the elective matrix being thus referred to as serum alcohol concentration (SAC) or plasma alcohol concentra-

tion (PAC). Driving under the influence, on the other hand, is a condition that, although not necessarily associated with fatal intoxication, results in cognitive impairment responsible for 25% of fatal road accidents, causing approximately 10,000 deaths *per* year in both Europe and the United States (2-5). The forensic determination of BAC is carried out by means of gas-chromatographic separation after headspace extraction (HS-GC) with either mass spectrometric (MSD) or flame ionization (FID) detection, using whole blood as the matrix of choice.

Operationally, beside AAI, cases a hospital and its clinical laboratory can be involved at various levels in the analysis of DUI, being in charge of sample collection alone or of the full confirmatory analysis. However, while the forensic analytical phase may be outsourced due to the need for specialized instrumentation and expertise, the resources of the preanalytical phase - from the personnel responsible for sample collection to the refrigerators used for sample storage - are often shared between diagnostic and forensic BAC testing. This has led to extensive discussions among laboratory specialists, which can be summarized as the need to identify, on one hand, the forensic defensibility characteristics of data produced by the diagnostic process and, on the other hand, the sustainability in a clinical environment of a specific process that guarantees the legal validity of the results (6-11).

To gain control over the preanalytical phase through operational procedures, it is necessary to understand the role and significance of various factors in determining BAC. Considering that the preanalytical phase in BAC accounts for 40% of the analytical result, compared to 20% for biological variability, one must have a comprehensive view of which factors are actually controllable (12). Notable contributions in this regard are provided by the scientific literature on forensic subjects (13). However, specifically for the clinical laboratory scientist, there is only one official document issued by the Clinical and laboratory standards institute (CLSI) (14). However, this document suffers from a series of limitations: it has never been revised since its initial publication in 1997, it is not currently officially available, and, most importantly, it only covers the preanalytical phase in three paragraphs (2.3 - 2.5) based on just six references.

The purpose of this literature review is therefore to provide the laboratory professional with the most comprehensive understanding of the factors determining the preanalytical phase of BAC, especially considering the coexistence of diagnostic and forensic processes in the clinical setting. A series of appendices to the text provides further insights into remarkable topics related with the BAC and the investigation thereof.

Literature search

Pubmed, Google Scholar and MEDLINE were searched for papers published until December 2022 with no restriction on language (see Appendix A). The search strategy based on a categorization of the subject according to a suitable model of the pre-analytics of a drug testing based on four major topics (namely, sampling, handling, contamination and matrix) is also represented in Figure 1. A typical query used for a preliminary search was as follows: ("blood alcohol" OR "blood ethanol") AND "sampl*" AND ("stabil*" OR "stor*" OR "temperature" OR "contaminat*" OR "factor*"). The search was then refined within each topic adjusting by the use of more selective terms (*e.g.*, "stabiliz*", "preserv*", "additive", "sealing", "leak*" for the handling factor or "haemol*", "clot*", "lipem*", "icter* for the matrix factor"). The literature search was further extended reviewing bibliography within each article issued before 1980s in order to retrieve any cited source that was eventually not electronically indexed (see Appendix A). Studies concerning *post-mortem* specimens or animals were excluded.

Pre-analytical factors

Sampling time

Ethanol is a very low molecular weight compound (46.07 g/mol and density of 0.789 g/mL at 20 °C) that can freely diffuse through the cell membrane (13). When ingested as a diluted water solution, it is absorbed *per* passive diffusion, first in the stom-

ach and then extensively in the intestine (13). The peak of blood absorption depends on the timing of gastric emptying but not on the concentration or the volume of the consumed beverage due to the very large absorptive surface of the proximal small bowel (15,16). Accordingly, the time to reach body fluids equilibrium is erratic and depends on both the stomach emptying and the variations of the blood flow of the gut (17).

In general, the fasting peak blood is observed within 45-60 minutes after finished drinking, however in some subjects it may already be present by as early as 15 minutes or delayed to up to 120 minutes from ingestion (13). In this phase, where EtOH is equilibrating between blood and tissues, there is a negative veno-arterous bias ($-\Delta_{V/A}$) up to - 0.20 g/L in the same tributary area (*e.g.*, cephalic vein *vs.* radial artery), and a positive veno-venous bias ($+\Delta_{V/V}$) between proximal and distal tributary areas (*e.g.*, up to 0.06 g/L forearm *vs.* opposite foot) (17-20). Venous BAC is also affected by large erratic fluctuations caused by the equilibration phenomena (21).

In the post-absorption phase that begins about 90 minutes post-drinking, the veno-arterous bias changes the sign ($+\Delta_{A/V}$) and no veno-venous bias ($\Delta_{V/V} \approx 0$) is usually observed (17-20). In this phase, the venous BAC reflects the distribution equilibrium between blood and tissues, as in the brain where it exerts its psychoactive action. In this phase, if blood is recollected within 10 minutes from the previous sample, the resampling bias lays

within the physiological random fluctuations (averaging ± 0.008 g/L and ± 0.010 g/L for venous and arterial blood, respectively) (13,22).

It should be noted that due to the kinetics of distribution to and from the tissues, the capillary blood shows an additive negative bias ranging by - 0.04 to - 0.06 g/L (17,23).

Sampling site

The superficial veins laying in the antecubital fossa of the forearm represent the elective sites of routine phlebotomy, however they are characterized by remarkable topographical variability (both within- and between-subject) and more anastomoses than the deeper arteries (24). Thus, even when $\Delta_{V/V} \approx 0$, it can be observed with fairly large prevalence (almost 50%) an inter-arm bias ($\Delta_{L/R}$) that ranges between 0.008 g/L and 0.024 g/L with significant between-subject variability (coefficient of variation, CV≈60%) (12,17,25). Despite the size of this bias is almost the same before and after the absorption peak, its fluctuations are much larger in the absorption phase (17). Remarkably, this bias has unpredictable direction as it is unrelated with handedness, gender, pattern of the superficial veins, level of the blood gasses and haematocrit (12,17,25).

Sampling technique

Since water content of blood and tissues shows very small within-subject variability (CV < 3.0%), any mechanical factor disturbing the perfusion flow can affect the equilibration of EtOH and in turn the BAC (22,26,27). For instance, a difficult positioning of an indwelling sampling device (catheter, butterfly needle) can increase up to five-fold the time-independent random BAC differences observed in the post-absorption phase (22).

The vasoconstriction of the tributary area induced by cooling (*i.e.*, at 13-15 °C) doubles both the size and the duration of the $\Delta_{A/V}$ when the EtOH is administered by infusion, whereas the vasodilation (*i.e.*, warming at 60 °C) halves only the size of the $\Delta_{A/V}$ (28). However, when EtOH is ingested, the duration (1 to 5 minutes) and the pressure applied to induce the venous stasis (60 to 100 mmHg) as well as the local ice-cooling of the skin do not have an effect over the BAC, while the body positioning (orthostatic reaction) and the physical activity produce a negative bias (29). This apparent contradiction between infused and ingested EtOH may eventually depend on the fact that the post-absorption route of distribution plays an unaddressed experimental and physiological role.

Chemical contamination

Ethanol volatizes quickly from surfaces (the well-known "cooling effect" experienced after skin cleansing) with half-life on skin of 11.7 second, so that theoretically is necessary to expose 1000 cm^2 of skin to 70% v/v alcoholic solution to achieve a BAC of 0.06 g/L (30).

Experimentally, a chemical contamination during sampling requires that the needle is suctioning while it is in direct contact with the alcoholic antiseptic (*e.g.*, pressing the soaked swab to stop bleeding while withdrawing the needle) (31-33). Otherwise, (*e.g.*, without pressing the soaked swab on the site of phlebotomy), with pre-evacuated tubes there is no contamination when the excess antiseptic is used (*i.e.*, 2 mL) and alcohol is not allowed to dry off (*i.e.*, 5 seconds waiting) (34). Likewise, the use of a syringe and thus of a controlled suction makes actually difficult to contaminate deliberately the specimens even pouring the alcohol directly onto the skin and inserting the needle shortly afterwards (35). Hence, if the correct amount of antiseptic is used (*i.e.*, 1 mL) and it is allowed to dry off (*i.e.*, 1 minute), contamination has only 5% probability or less to happen (36,37). However, the spurious BAC due to a chemical contamination is erratic and unpredictable (*e.g.*, depending on the degree of swab squeezing), and values ranging from 0.005 g/L to up to 6.0 g/L have been reported (33,35,36).

It must be noted that when the experimentation involves inebriated subjects, the EtOH swabbing seems to give an average increase of 0.05 g/L of the BAC regardless of whether a pre-evacuated tube or a syringe was used (38-40). As already discussed for the mode of blood sampling, the dis-

crepancy that occurs when the study involves inebriated subjects may rather be the effect of some uncontrolled experimental factor.

The effect of the chemical contamination depends on the analytical method used for measuring the BAC when skin disinfection is made with alcohol like isopropyl or amyl, since the spectrophotometric enzymatic assay based on yeast alcohol dehydrogenase (ADH) is highly selective for EtOH (37,41-43). Indeed, contaminating with isopropanol (*i.e.*, 2-propanol) gives a negative bias when the BAC quantitation *via* the HS-GC uses this alcohol in place of the isomer n-propanol (1-propanol) as the internal standard (44).

Microbial contamination

Even though the pre-evacuated tube used for blood sampling is a closed and sterile system that avoids environmental contamination during and after the blood withdrawal, the incorrect skin disinfection can be the source of microbes' contamination during phlebotomy. Indeed, *Proteobacteria*, *Bacterotides* and *Staphylococcaceae* colonize the moisty skin of the antecubital fossa (45,46). Among the fungi, *Malassezia* predominates on *Candida* (47). All such microbes are able to synthesize anaerobically the EtOH *via* the fermentation (with production of small amounts of ethyl acetate as by-product), except for *Malassezia* that can only hydrolyse the fatty acids ethyl esters (48,49).

In a freshly collected blood sample, the contaminating microbes are in the disadvantage respect to the far more numerous erythrocytes for the uptake of glucose. Therefore, in such unfavourable conditions of substrates availability, the microbial ADH operates for the salvage pathway by reducing EtOH to restore nicotinamide adenine dinucleotide NAD^+ from NADH + H^+ (50-52). As the consequence, the bias of the BAC from a contaminated specimen is expected to be negative, except for the case of supplementation of the sample with extra-glucose as in a banked blood bag (53,54). Accordingly, the BAC of an EtOH-free (or below the detection limit) specimen remains unchanged regardless such factor as the use of sodium fluoride (NaF) preservative, the duration and temperature of the storage and the conditions of handling (*e.g.*, heat exposure, repeated uncapping and sampling) (55-60). Likewise, the environmental contamination of the blood matrix during the sampling and processing operations (*e.g.*, uncapping, decanting, pipetting) is unlikely to occur in a typical laboratory setting even when the sample is forcedly exposed or handled carelessly (57).

The NaF represents the major anti-microbial agent used for stabilizing the BAC and it has been recommended since very early in the scientific literature (61). At a concentration ranging within 200 mM and 300 mM (≈1% w/v), it causes the cell death within 24 hours, while at lower concentrations (*e.g.*, 100 mM) it can take up to 72 hours (62). Hence, at least within 24 hours the preservation of BAC is NaF-independent (51). Nonetheless, such microbes as *Pseudomonas* and *Serratia* can survive to NaF exposure up to 600 mM (≈2% w/v) (52). Hence, NaF acts as a second-line defense against contamination but is not superior to the primary sterility of the collection tube. Remarkably, in an ordinary tube used for glucose testing, the concentration of NaF is 60 mM (*i.e.*, 0.25% w/v or 2.5 mg/mL whole blood) and thus it is inadequate for preventing the growth of the microbes.

Sample container

Ethanol is an organic compound whose octanol/water partition coefficient ($logP_{o/w}$) lays close to -0.3, therefore it cannot be adsorbed into the gel separator of clinical chemistry tubes as it requires $logP_{o/w} > 3$ (63-66). Accordingly, no adsorption bias is observed when the serum BAC is measured in plain or gel separator tubes (67,68). On the contrary, as a volatile organic solvent, the EtOH permeates the walls of tube made of single-layered polypropylene, but not the inner-layered polyethylene terephthalate and the glass borosilicate (64,69-71). In general, the permeation bias between a double-walled plastic and a glass tube rests within 1% of BAC regardless of the storage temperature (69,72). This condition is also unaffected by the extension of the contact surface of blood on the tube walls (*i.e.*, whether the tube is stored standing up or laying down) (70).

Occasionally, the gel separator can be the source of some chemical contamination, especially in the past due to the solvents used for the manufacturing of the rubber stoppers (unidentified substance) and gel separator of serum tubes (toluene, 1-butanol, ethylbenzene and xylene) (73-75). This was also found in whole blood sodium fluoride/oxalate (NaF/Ox) tubes containing traces of isobutylene (76). However, none of the above compound interferes with the spectrophotometric enzymatic assay of BAC. Conversely, chemical contaminants released from gel separator tubes and able to interfere with the HS-GC analysis carried out without the mass selective detector (MSD) were also reported recently. Particularly, these were a 1-propanol-like substance and ethyl chloride (chloroethane, C_2H_5Cl), both capable to positively bias the BAC as almost co-eluting with EtOH (77-79).

Dispersion

The complete solubility of EtOH into water is due to the presence of the hydroxyl moiety on the very short carbon chain, so that the water/air partition coefficient ($logP_{w/a}$) measured in distilled water does not significantly differ from whole blood and plasma ($logP_{w/a}\approx3.33$, $logP_{w/a}\approx3.24$ and $logP_{w/a}\approx3.31$, respectively) (80,81). Consequently, the evaporation bias of an uncapped serum or heparinized plasma tube resting at 22.1 - 25.1 °C temperature and 55% maximum air humidity averages - 3.0% within 30 minutes, - 5.0% after 1 hour and reaches a maximum of - 10.0% after 3 hours (82).

In a capped tube, EtOH collects into the air volume above the specimen (so-called "headspace"), reaching a concentration (*i.e.*, partial pressure) that only depends on the temperature as explained by the Henry-Dalton's law. Accordingly, assuming that the distribution volume of EtOH contained in a 0.5 mL sample is about 2500 mL airspace at room temperature (20 °C), and that the headspace in a gas-tight sealed tube is about 1 mL when correctly filled, then the bias produced by venting once the headspace is less than -3% even at 40 °C (80,81,83,84). Therefore, a leak from the sealed cap occurring during the storage requires an extremely long period of time (*e.g.*, 1 year) to reach a bias of - 1% (85).

As the loss of EtOH due to the partition in the headspace is negligible, there is no bias between 6 mL and 10 mL volume NaF/Ox pre-evacuated sterile tubes (*i.e.*, different sample-to-headspace ratio) when these are correctly filled with the whole blood (86). However, when the NaF/Ox tube is partially filled so that the NaF reaches a final concentration of 2% or 5% w/v, the bias produced at room temperature by the headspace vent is - 3.0% and - 9.0%, respectively (87). This bias is due to the "salting-out" effect of the concentrated NaF in the matrix that increases the EtOH evaporation into the headspace (81).

Remarkably, for those longer chain alcohols (*e.g.*, propyl and butyl) used as the analytical internal standard in HS-GC, the salting-out is stronger because of their naturally lower water solubility (88,89). As a consequence, in case of partial filling the EtOH/internal standard peak area ratio in the sample obtained from 1% w/v NaF/Ox tube is fictitiously reduced giving rise to bias up to - 3% if no salt-saturated sample preparation is adopted (87,90-92).

Sample storage

In healthy subjects, it can be found no ADH in a significant concentration either within erythrocytes or free in serum (93). However, the rate of acetaldehyde formation in ethylene diamine-tetraacetic acid-anticoagulated (EDTA) whole blood containing EtOH steeply increases regardless of the addition of inhibitors (*e.g.*, citric acid, iodoacetic acid, fluoride, aminoatriazole, azide and pyrazole) of oxidases that may attack very short-chained aliphatic compounds (*e.g.*, glyceraldehyde-3-phosphate dehydrogenase, catalase, ADH) (94). Therefore, a non-enzymatic oxidation (NEO) of EtOH exists and its kinetic depends on storage temperature, matrix oxygenation and haemoglobin concentration (see Appendix B) (94-97).

Based on studies whose conditions are compatible with sampling blood from a patient using a pre-evacuated sterile tube (*i.e.*, blood suctioned within by the vacuum), a whole blood sample can be delivered at room temperature and stored refrigerated (*i.e.*, - 20 °C) for up to 14 days with no forensi-

cally and diagnostically significant bias (see Appendix C) (55,57,58,69,72,98). However, in case of harsh handling of the specimen before storage or analysis (*e.g.*, exposure to elevated room temperature, high thermic excursion or external transportation without refrigeration), the bias increases up to - 0.20 g/L yet within the third day from collection (56). The prolonged heating surely enhances the kinetics of the NEO, but it is also likely to cause a loosening of the gas-tight cap through which allows the venting of the tube. In fact, even a short (< 20 min) but extreme heating (burning) gives negligibly biased of the BAC if the collection tube remains intact (99).

The negative bias tends to appear by 15 days up to 3 months depending on how much the temperature, oxygenation and time are actually affecting the kinetics of NEO (see Appendix C) (95). For instance, after 1 month at - 20 °C the bias averages - 6% when the headspace is 20% (*i.e.*, 1/5 of the tube volume) (100). However, venting the tube at least once before or during storage almost doubles the bias regardless of both the temperature (by - 10 °C to + 25 °C) and the duration (days to years) of storage (55,58,60,101,102). In general, a mild storage condition (*i.e.*, + 4 °C) gives a bias that is on average within - 0.04 g/L for one-year and - 0.20 g/L for three years (103,104). As the size of the storage bias correlates well with the storage time, the correlation between storage bias and the original BAC depends on the length of the time interval over which it is computed (*e.g.*, r = 0.80 up to 6 months, r = 0.23 up to 13 months *vs*. and r < 0.17 more than 5 years) (55,58,60,100).

Despite the lack of haemoglobin, the stored plasma shows - 7.8% bias after 2 months and up to - 25.2% after 5 months at - 20 °C (105). This bias is reduced to - 2% after 6 months if the sample is deproteinized by acid precipitation (106). Maybe, the loss of EtOH from plasma depends on the effect of residual iron within the matrix (also freed by the acid precipitation), as well as on the integrity of the gas-tight seal at low temperature.

Finally, an effectively preserving activity of BAC is observed in carbon monoxide-saturated whole blood where no significant bias occurs due to the blockage of oxyhaemoglobin formation (107). Comparable result can be achieved by deoxygenating the whole blood with bubbling nitrogen (94). This mechanism can be relevant for preserving the BAC in the samples used for quality control and proficiency testing, and it suggests that the smoking condition might act as a preanalytical factor interacting with other factors in the bias associated with the conditions of storage (55).

Matrix integrity

The spectrophotometric enzymatic measurement of BAC in serum/plasma is based on the UV absorption peak of NADH + H^+ at 340 nm, that is produced by the ADH according to a 1:1 stoichiometry during the oxidation of EtOH (108). This absorption peak is close to that produced by free oxygenated haemoglobin. However, the modern enzymatic assays use a 20-fold dilution to measure the BAC, so that no interference (*i.e.*, positive bias) is expected even with gross haemolysis (0.8 g/dL free haemoglobin) (109). It is likely that the haemolysis bias up to - 10% is caused by the oxidation of EtOH occurring before the enzymatic analysis, for instance as the consequence of the release of catalase from the erythrocytes (109-111). As this also accelerates the depletion of the cellular antioxidants that antagonize the NEO (see Appendix B), storing a sample with haemolysis causes a bias up to - 0.03 g/L within one week (112,113).

Ethanol can bind to hydrophobic sites of the albumin *via* the methyl group (114). Although this interaction can displace drugs like diazepam, warfarin and corticosteroids, the affinity constant toward albumin is actually large (K_d = 53.1 ± 3.1 mM or ≈2.4 g/L) (115-117). Since the formation of clots causes a minimal loss of protein and there is no sizable bound fraction of EtOH, the serum/plasma ratio of BAC is almost unity with very low variability (CV 0.01% to 0.03%) (72,118). However, since both the size and the water content (by 40% to 80% w/v) of clots is highly variable, the formation of clots within a whole blood specimen alters the distribution of water and makes unpredictable the bias caused by the inhomogeneity of the BAC (119,120). Thus, a specimen of whole blood with

clots should be homogenized before the determination of BAC but the negative bias arising from the grinding process averages - 0.01 g/L (121).

Other relevant factors related to the control of BAC

Biosynthesis

In liver and kidney of humans, the mitochondrial oxidative decarboxylation of pyruvate produces acetaldehyde (122). Under hypoxic conditions, acetaldehyde is metabolized to EtOH by the cytosolic ADH in a redox reaction that converts the cofactor nicotinamide adenine dinucleotide from reduced ($NADH + H^+$) to oxidized (NAD^+) form (122). The frequency distribution of the BAC produced by this endogenous synthesis in ostensibly healthy and sober individuals is strongly left-skewed, with the 95^{th} and 99.4^{th} percentiles (age-, gender- and race-unrelated) corresponding to 0.012 g/L and 0.020 g/L respectively (123,124). To date, no individual value above 0.040 g/L has been reported (125).

Just in case of a rare clinical condition known as "auto-brewery" or "gut fermentation" syndrome (no more than 20 cases correctly identified in literature until 2021), usually associated with severe fungal infections, gastric resection or Crohn disease, the BAC can be as high as 0.5 g/L and up to 2.0 g/L under carbohydrates loading (126,127).

Biochemistry

Because of the negligible binding to proteins and membranes, the distribution of EtOH follows the water content of blood components and thus it is partitioned between the serum/plasma fraction and the cytosol of the erythrocytes (with minor contributions from platelets and leukocytes) (128). Since the average water content of erythrocytes is ≈75% w/v (weight/volume), ≈85% w/v of whole blood and ≈95% w/v of serum/plasma, then the serum/whole blood as well as plasma/whole blood concentration ratio reflects the average water content of the blood components weighted by their relative abundance and is equal to 1.10-1.14 (26,72,118,128,129). This ratio varies more between- than within-subjects, and the CV < 3% is less than the biological variability of the haematocrit (CV≈5.5%) (26,130). Accordingly, the BAC measured in serum/plasma is unaffected by the haematocrit unless there is an extreme shift as in the polycythaemia or severe anaemia as during haemorrhage (73,118,128,129,131).

As water content rules the BAC, sizable deviations from the ratio of 1.10-1.14 can arise for a change in proteins and lipids content of serum/plasma (129,132). In this regard, only extreme alimentary, stress-related or genetic hyperlipidaemias can eventually reduce water content of serum by no more than 5%. On the contrary, much larger effect (in either directions) is expected for hyper- or hypoproteinaemia *via* the strong regulation of the oncotic pressure, that in turn affects the distribution of EtOH to the tissues (especially the skeletal muscle).

Discussion

For the clinical laboratory, the BAC represents a significant organizational and cultural challenge due to its dual clinical and forensic significance. Even when not directly involved in the analysis of DUI cases - often outsourced to external laboratories for the need for specific instrumentation and expertise - the clinical laboratory may be involved in managing cases that, due to the circumstances in which they originated, can assume (often unpredictably) medico-legal relevance. It is here that the need for defensibility arises, that is, the ability to justify the forensic validity of the data, not as an added value but as an integral part of the diagnostic process. To this concern, it must be not disregarded the remarkable divide of magnitude between the allowable total error (TEa) set for the diagnostic BAC (between 20% and 9% of Clinical laboratory improvement amendments (CLIA) and Guidelines of the German Federal Medical Council (Rili-BAEK), respectively) and the combined measurement uncertainty recommended for the forensic BAC (within 4%) (133-138).

The preanalytical factors of the BAC (Figure 1) can generally be distinguished into two types: those related to the behaviour of EtOH within the patient's body, which instantaneously determine the

BAC at the time of phlebotomy, and those related to the behaviour of EtOH within the blood specimen, which determine the resulting BAC after phlebotomy. Remarkably, the factors of the first type (*i.e.*, time and site of sampling) have a limitedly controllable nature *via* the operative procedures (as in the case of a polytraumatized patient involved in a car accident whose exact time of last consumption is unknown). Notwithstanding their consideration falls within the logic of defensibility of the data, that for instance can be achieved through the postanalytical phase by building up an uncertainty budget derived from the knowledge of the associated bias and imprecision. The same applies to EtOH biosynthesis and biochemistry, the postanalytical control of which is perhaps the most significant expression of the dual clinical and forensic significance of the BAC, where a possible reference interval of endogenous EtOH forms the basis for a more realistic discussion on how to deal analytically with a zero-tolerance policy for EtOH consumption.

The factors of the second group (namely from sampling technique to matrix integrity), on the other hand, are operationally controllable. Therefore, they are the subject of more intense debate when the issue of coexistence between clinical and forensic preanalytics of the BAC is raised, because they determine the degree to which the operations need to be complicated to ensure the adequacy of the preanalytical process. In this regard, it should be noted that the preanalytics of the BAC is perhaps the most studied among laboratory tests and spans nearly a century of scientific publications (see the Appendix A). On the one hand, this is a favourable aspect because it indicates extensive characterization of the subject matter. On the other hand, it must be recognized that precisely because of this extensive characterization, the produced evidence has stratified through the evolution of knowledge and means that have accompanied the recent history of clinical chemistry. In other words, retrospective analysis often reveals discrepancies or incongruences among the evidence, making it difficult to grasp the correct indications provided by direct investigation or observation. In particular, this refers to the age-old question of whether it is necessary to collect and preserve the blood sample for forensic BAC separately from diagnostic BAC, that is, whether the preanalytical processes must necessarily be distinct.

In the investigation of preanalytics, a crucial aspect is represented by the relationship established between the experimental design and the factors effectively controlled by the experimenter (139). When considering the vast amount of literature on the preanalytics of the BAC, it is possible to explain and understand this only by taking into account the impact produced by the introduction of new blood collection devices, sterile and pre-evacuated, which have been increasingly used since the 1970s. These devices are the key that translates the control of preanalytical factors, as characterized by Smalldon, Brown and colleagues in their seminal works (even though they did not use such devices), into practice (95,96). If their experimental role is not explicitly taken into consideration (which is easy because they become part of routine activity for sample collection in studies), their presence or absence in a preanalytical study acts as a confounder (see Appendix C). Supporting this consideration, it is worth noting that no preanalytics study of the BAC has ever compared the use of these new devices with the previous technique based on blood collection with a syringe and dispensing into various pre-added but open tubes.

Given the above, there is no evidence to support the need for intensifying and differentiating the use of devices and procedures for AAI or DUI cases to mitigate bias for the latter if pre-evacuated and sterile collection tubes are used according to the correct phlebotomy procedures (*i.e.*, complete filling and mixing, vacuum sealing, and no haemolysis) (see Table 1). Indeed, under such conditions, the bias arising from intra- and extra-mural transport conditions, processing (*e.g.*, at room temperature for < 3 hours), short-term storage (*e.g.*, at + 4 °C for < 3 months), and long-term storage (*e.g.*, at - 20 °C for < 3 years) adopted for diagnostics is compatible with the requirements for forensic BAC. This overall agrees with the conclusions reached by other authors through different paths, that there is no requirement for a specific preser-

Table 1. The "seven pillars" of a diagnostically reliable and forensically defensible unified preanalytics of blood alcohol concentration

The "seven pillars" of a diagnostically reliable and forensically defensible unified preanalytics of BAC
1. use non-alcoholic antiseptics for skin cleansing
2. change the site of phlebotomy instead of probing/manipulating
3. choose either heparin or EDTA anticoagulant
4. use only pre-evacuated sterile collection tubes of the smallest capacity (*i.e.*, 2.5 mL)
5. avoid air suction, tube venting and haemolysis before storage
6. freeze only for monthly or yearly storage
7. beware of vented or opened tubes especially for reanalysis

BAC - blood alcohol concentration. EDTA - ethylene diamine-tetra-acetic acid.

vative for forensic BAC, such as NaF (which, moreover, introduces additional complications like salting-out or haemolysis) (140,141).

The only necessary special care relates to the use of non-alcoholic disinfectants for forensic BAC, which also benefits the diagnostics of AAI. In this regard, it should be remarked that guidelines provided by official laboratory medicine bodies such as the European federation of clinical chemistry and laboratory medicine (EFLM) do not ban alcoholic antiseptics for BAC request, but advice to take adequate time for drying off the alcohol before the venipuncture (142). Therefore, the present recommendation should be regarded as a reinforcement of that guidance, in a way that makes the preparation for blood collection for BAC analysis 'error-proof' against the pressures that a challenging operational context (*e.g.*, the Emergency Room) may exert on the correct timing of the procedure.

Since these disinfectants offer the same safety and usage procedures as alcoholic ones and are widely available in the market without significant budget impact, their substitutionary use (if not already in place) can be safely adopted for all blood sampling procedures (24). Therefore, the only necessary differentiation remains the adoption of a chain of custody for sample traceability whenever appropriate for the explicit medico-legal end of the requested BAC.

Therefore, it can be concluded that the execution of correct diagnostic preanalytical procedures guarantees both the diagnostic safety and forensic defensibility of clinical data, and no evidence contrary to a common preanalytical process can be found in the literature. It must be remarked that any deviation from these conditions magnifies the bias guaranteed by the use of pre-evacuated sterile devices (*e.g.*, opening of the tube for matrix sampling), compromising the integrity of the BAC, both forensic and diagnostic, regardless of the procedures adopted before and after its reanalysis.

Acknowledgments

The author wishes to thank Dr. Med. U. Wantoch von Rekowski for helping with the German papers, and Dr. Med M. Sapio for suggestions. This work is dedicated to the memory of A.M. Lostia, PhD, clever researcher and friend.

Potential conflict of interest

None declared.

References

1. *World Health Organization (WHO). Global status report on alcohol and health 2018. Geneve (CH): World Health Organization; 2018. Available from https://www.who.int/publications/i/item/9789241565639. Accessed May 5th 2023.*
2. *Griswold MG, Fullman N, Hawley C, Arian N, Zimsen SRM, Tymeson HD, et al. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2018;392:1015-35. https://doi.org/10.1016/S0140-6736(18)31310-2*

Observational and interventional study design types; an overview

Matthew S. Thiese

Rocky Mountain Center for Occupational and Environmental Health, University of Utah, Salt Lake City, Utah, USA

*Corresponding author: Matt.thiese@hsc.utah.edu

Abstract

The appropriate choice in study design is essential for the successful execution of biomedical and public health research. There are many study designs to choose from within two broad categories of observational and interventional studies. Each design has its own strengths and weaknesses, and the need to understand these limitations is necessary to arrive at correct study conclusions.

Observational study designs, also called epidemiologic study designs, are often retrospective and are used to assess potential causation in exposure-outcome relationships and therefore influence preventive methods. Observational study designs include ecological designs, cross sectional, case-control, case-crossover, retrospective and prospective cohorts. An important subset of observational studies is diagnostic study designs, which evaluate the accuracy of diagnostic procedures and tests as compared to other diagnostic measures. These include diagnostic accuracy designs, diagnostic cohort designs, and diagnostic randomized controlled trials.

Interventional studies are often prospective and are specifically tailored to evaluate direct impacts of treatment or preventive measures on disease. Each study design has specific outcome measures that rely on the type and quality of data utilized. Additionally, each study design has potential limitations that are more severe and need to be addressed in the design phase of the study. This manuscript is meant to provide an overview of study design types, strengths and weaknesses of common observational and interventional study designs.

Key words: study design; epidemiology; observational study; randomized trials; study strengths and weaknesses

Received: October 14, 2013 *Accepted: April 12, 2014*

Introduction

Study design plays an important role in the quality, execution, and interpretation of biomedical and public health research (1-12). Each study design has their own inherent strengths and weaknesses, and there can be a general hierarchy in study designs, however, any hierarchy cannot be applied uniformly across study design types (3,5,6,9). Epidemiological and interventional research studies include three elements; 1) definition and measure of exposure in two or more groups, 2) measure of health outcome(s) in these same groups, and 3) statistical comparison made between groups to assess potential relationships between the exposure and outcome, all of which are defined by the researcher (1-4,8,13). The measure of exposure in epidemiologic studies may be tobacco use ("Yes" *vs*. "No") to define the two groups and may be the treatment (Active drug *vs*. placebo) in interventional studies. Health outcome(s) can be the development of a disease or symptom (e.g. lung cancer) or curing a disease or symptom (e.g. reduction of pain). Descriptive studies, which are not epidemio-

logical or interventional, lack one or more of these elements and have limited application. High quality epidemiological and interventional studies contain detailed information on the design, execution and interpretation of results, with methodology clearly written and able to be reproduced by other researchers.

Research is generally considered as primary or secondary research. Primary research relies upon data gathered from original research expressly for that purpose (1,3,5). Secondary research focuses on single or multiple data sources that are not collected for a single research purpose (14,15). Secondary research includes meta-analyses and best practice guidelines for treatments. This paper will focus on the study designs and their strengths, weaknesses, and common statistical outcomes of primary research.

The choice of a study design hinges on many factors, including prior research, availability of study participants, funding, and time constraints. One common decision point is the desire to suggest causation. The most common causation criteria are proposed by Hill (16). Of these, demonstrating temporality is the only mandatory criterion for suggesting temporality. Therefore, prospective studies that follow study participants forward through time, including prospective cohort studies and interventional studies, are best suited for suggesting causation. Causal conclusions cannot be proven from an observational study. Additionally, causation between an exposure and an outcome cannot be proven by one study alone; multiple studies across different populations should be considered when making causation assessments (17).

Primary research has been categorized in different ways. Common categorization schema include temporal nature of the study design (retrospective or prospective), usability of the study results (basic or applied), investigative purpose (descriptive or analytical), purpose (prevention, diagnosis or treatment), or role of the investigator (observational or interventional). This manuscript categorizes study designs by observational and interventional criteria, however, other categorization methods are described as well.

Observational and interventional studies

Within primary research there are observational studies and interventional studies. Observational studies, also called epidemiological studies, are those where the investigator is not acting upon study participants, but instead observing natural relationships between factors and outcomes. Diagnostic studies are classified as observational studies, but are a unique category and will be discussed independently. Interventional studies, also called experimental studies, are those where the researcher intercedes as part of the study design. Additionally, study designs may be classified by the role that time plays in the data collection, either retrospective or prospective. Retrospective studies are those where data are collected from the past, either through records created at that time or by asking participants to remember their exposures or outcomes. Retrospective studies cannot demonstrate temporality as easily and are more prone to different biases, particularly recall bias. Prospective studies follow participants forward through time, collecting data in the process. Prospective studies are less prone to some types of bias and can more easily demonstrate that the exposure preceded the disease, thereby more strongly suggesting causation. Table 1 describes the broad categories of observational studies: the disease measures applicable to each, the appropriate measures of risk, and temporality of each study design. Epidemiologic measures include point prevalence, the proportion of participants with disease at a given point in time, period prevalence, the proportion of participants with disease within a specified time frame, and incidence, the accumulation of new cases over time. Measures of risk are generally categorized into two categories: those that only demonstrate an association, such as an odds ratio (and some other measures), and those that demonstrate temporality and therefore suggest causation, such as hazard ratio. Table 2 outlines the strengths and weaknesses of each observational study design.

Observational studies

Ecological study design

The most basic observational study is an ecological study. This study design compares clusters of people, usually grouped based on their geographical location or temporal associations (1,2,6,9). Ecological studies assign one exposure level for each distinct group and can provide a rough estimation of prevalence of disease within a population. Ecological studies are generally retrospective. An example of an ecological study is the comparison of the prevalence of obesity in the United States and France. The geographic area is considered the exposure and the outcome is obesity. There are inherent potential weaknesses with this approach, including loss of data resolution and potential misclassification (10,11,13,18,19). This type of study design also has additional weaknesses. Typically these studies derive their data from large databases that are created for purposes other than research, which may introduce error or misclassification (10,11). Quantification of both the number of cases and the total population can be difficult, leading to error or bias. Lastly, due to the limited amount of data available, it is difficult to control for other factors that may mask or falsely suggest a relationship between the exposure and the outcome. However, ecological studies are generally very cost effective and are a starting point for hypothesis generation.

Proportional mortality ratio study design

Proportional mortality ratio studies (PMR) utilize the defined well recorded outcome of death and subsequent records that are maintained regarding the decedent (1,6,8,20). By using records, this study design is able to identify potential relationships between exposures, such as geographic location, occupation, or age and cause of death. The epidemiological outcomes of this study design are proportional mortality ratio and standardized mortality ratio. In general these are the ratio of the proportion of cause-specific deaths out of all deaths between exposure categories (20). As an example, these studies can address questions about higher proportion of cardiovascular deaths among different ethnic and racial groups (21). A significant drawback to the PMR study design is that these studies are limited to death as an outcome (3,5,22). Additionally, the reliance on death records makes it difficult to control for individual confounding factors, variables that either conceal or falsely demonstrate associations between the exposure and outcome. An example of a confounder is tobacco use confounding the relationship between coffee intake and cardiovascular disease. Historically people often smoked and drank coffee while on coffee breaks. If researchers ignore smoking they would inaccurately find a strong relationship between coffee use and cardiovascular disease, where some of the risk is actually due to smoking. There are also concerns regarding the accuracy of death certificate data. Strengths of the study design include the well-defined outcome of death, the relative ease and low cost of obtaining data, and the uniformity of collection of these data across different geographical areas.

Cross-sectional study design

Cross-sectional studies are also called prevalence studies because one of the main measures available is study population prevalence (1-12). These studies consist of assessing a population, as represented by the study sample, at a single point in time. A common cross-sectional study type is the diagnostic accuracy study, which is discussed later. Cross-sectional study samples are selected based on their exposure status, without regard for their outcome status. Outcome status is obtained after participants are enrolled. Ideally, a wider distribution of exposure will allow for a higher likelihood of finding an association between the exposure and outcome if one exists (1-3,5,8). Cross-sectional studies are retrospective in nature. An example of a cross-sectional study would be enrolling participants who are either current smokers or never smokers, and assessing whether or not they have respiratory deficiencies. Random sampling of the population being assessed is more important in cross-sectional studies as compared to other observational study designs. Selection bias from non-random sampling may result in flawed measure of prevalence and calculation of risk. The study sample is assessed for both exposure and outcome at a single point in time. Because both exposure and

outcome are assessed at the same time, temporality cannot be demonstrated, i.e. it cannot be demonstrated that the exposure preceded the disease (1-3,5,8). Point prevalence and period prevalence can be calculated in cross-sectional studies. Measures of risk for the exposure-outcome relationship that can be calculated in cross-sectional study design are odds ratio, prevalence odds ratio, prevalence ratio, and prevalence difference. Cross-sectional studies are relatively inexpensive and have data collected on an individual which allows for more complete control for confounding. Additionally, cross-sectional studies allow for multiple outcomes to be assessed simultaneously.

Case-control study design

Case-control studies were traditionally referred to as retrospective studies, due to the nature of the study design and execution (1-12,23,24). In this study design, researchers identify study participants based on their case status, i.e. diseased or not diseased. Quantification of the number of individuals among the cases and the controls who are exposed allow for statistical associations between exposure and outcomes to be established (1-3,5,8). An example of a case control study is analysing the relationship between obesity and knee replacement surgery. Cases are participants who have had knee surgery, and controls are a random sampling of those who have not, and the comparison is the relative odds of being obese if you have knee surgery as compared to those that do not. Matching on one or more potential confounders allows for minimization of those factors as potential confounders in the exposure-outcome relationship (1-3,5,8). Additionally, case-control studies are at increased risk for bias, particularly recall bias, due to the known case status of study participants (1-3,5,8). Other points of consideration that have specific weight in case-control studies include the appropriate selection of controls that balance generalizability and minimize bias, the minimization of survivor bias, and the potential for length time bias (25). The largest strength of case-control studies is that this study design is the most efficient study design for rare diseases. Additional strengths include low cost, relatively fast execution compared to cohort studies, the ability to collect individual participant specific data, the ability to control for multiple confounders, and the ability to assess multiple exposures of interest. The measure of risk that is calculated in case-control studies is the odds ratio, which are the odds of having the exposure if you have the disease. Other measures of risk are not applicable to case-control studies. Any measure of prevalence and associated measures, such as prevalence odds ratio, in a case-control study is artificial because the researcher arbitrarily sets the proportion of cases to non-cases in this study design. Temporality can be suggested, however, it is rarely definitively demonstrated because it is unknown if the development of the disease truly preceded the exposure. It should be noted that for certain outcomes, particularly death, the criteria for demonstrating temporality in that specific exposure-outcome relationship are met and the use of relative risk as a measure of risk may be justified.

Case-crossover study design

A case-crossover study relies upon an individual to act as their own control for comparison issues, thereby minimizing some potential confounders (1,5,12). This study design should not be confused with a crossover study design which is an interventional study type and is described below. For case-crossover studies, cases are assessed for their exposure status immediately prior to the time they became a case, and then compared to their own exposure at a prior point where they didn't become a case. The selection of the prior point for comparison issues is often chosen at random or relies upon a mean measure of exposure over time. Case-crossover studies are always retrospective. An example of a case-crossover study would be evaluating the exposure of talking on a cell phone and being involved in an automobile crash. Cases are drivers involved in a crash and the comparison is that same driver at a random timeframe where they were not involved in a crash. These types of studies are particularly good for exposure-outcome relationships where the outcome is acute and well defined, e.g. electrocutions, lacerations, automobile crashes, etc. (1,5). Exposure-outcome

relationships that are assessed using case-crossover designs should have health outcomes that do not have a subclinical or undiagnosed period prior to becoming a "case" in the study (12). The exposure is cell phone use during the exposure periods, both before the crash and during the control period. Additionally, the reliance upon prior exposure time requires that the exposure not have an additive or cumulative effect over time (1,5). Case-crossover study designs are at higher risk for having recall bias as compared with other study designs (12). Study participants are more likely to remember an exposure prior to becoming a case, as compared to not becoming a case.

Retrospective and prospective cohort study design

Cohort studies involve identifying study participants based on their exposure status and either following them through time to identify which participants develop the outcome(s) of interest, or look back at data that were created in the past, prior to the development of the outcome. Prospective cohort studies are considered the gold standard of observational research (1-3,5,8,10,11). These studies begin with a cross-sectional study to categorize exposure and identify cases at baseline. Disease-free participants are then followed and cases are measured as they develop. Retrospective cohort studies also begin with a cross-sectional study to categorize exposure and identify cases. Exposures are then measured based on records created at that time. Additionally, in an ideal retrospective cohort, case status is also tracked using historical data that were created at that point in time. Occupational groups, particularly those that have regular surveillance or certifications such as Commercial Truck Drivers, are particularly well positioned for retrospective cohort studies because records of both exposure and outcome are created as part of commercial and regulatory purposes (8). These types of studies have the ability to demonstrate temporality and therefore identify true risk factors, not associated factors, as can be done in other types of studies.

Cohort studies are the only observational study that can calculate incidence, both cumulative incidence and an incidence rate (1,3,5,6,10,11). Also, because the inception of a cohort study is identical to a cross-sectional study, both point prevalence and period prevalence can–be calculated. There are many measures of risk that can be calculated from cohort study data. Again, the measures of risk for the exposure-outcome relationship that can be calculated in cross-sectional study design of odds ratio, prevalence odds ratio, prevalence ratio, and prevalence difference can be calculated in cohort studies as well. Measures of risk that leverage a cohort study's ability to calculate incidence include incidence rate ratio, relative risk, risk ratio, and hazard ratio. These measures that demonstrate temporality are considered stronger measures for demonstrating causation and identification of risk factors.

Diagnostic testing and evaluation study designs

A specific study design is the diagnostic accuracy study, which is often used as part of the clinical decision making process. Diagnostic accuracy study designs are those that compare a new diagnostic method with the current "gold standard" diagnostic procedure in a cross-section of both diseased and healthy study participants. Gold standard diagnostic procedures are the current best-practice for diagnosing a disease. An example is comparing a new rapid test for a cancer with the gold standard method of biopsy. There are many intricacies to diagnostic testing study designs that should be considered. The proper selection of the gold standard evaluation is important for defining the true measures of accuracy for the new diagnostic procedure. Evaluations of diagnostic test results should be blinded to the case status of the participant. Similar to the intention-to-treat concept discussed later in interventional studies, diagnostic tests have a procedure of analyses called intention to diagnose (ITD), where participants are analysed in the diagnostic category they were assigned, regardless of the process in which a diagnosis was obtained. Performing analyses according to an a priori defined protocol, called *per* protocol analyses (PP or PPA), is another potential strength to diagnostic study testing. Many measures of the new

diagnostic procedure, including accuracy, sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio can be calculated. These measures of the diagnostic test allow for comparison with other diagnostic tests and aid the clinician in determining which test to utilize.

Interventional study designs

Interventional study designs, also called experimental study designs, are those where the researcher intervenes at some point throughout the study. The most common and strongest interventional study design is a randomized controlled trial, however, there are other interventional study designs, including pre-post study design, non-randomized controlled trials, and quasi-experiments (1,5,13). Experimental studies are used to evaluate study questions related to either therapeutic agents or prevention. Therapeutic agents can include prophylactic agents, treatments, surgical approaches, or diagnostic tests. Prevention can include changes to protective equipment, engineering controls, management, policy or any element that should be evaluated as to a potential cause of disease or injury.

Pre-post study design

A pre-post study measures the occurrence of an outcome before and again after a particular intervention is implemented. A good example is comparing deaths from motor vehicle crashes before and after the enforcement of a seat-belt law. Pre-post studies may be single arm, one group measured before the intervention and again after the intervention, or multiple arms, where there is a comparison between groups. Often there is an arm where there is no intervention. The no-intervention arm acts as the control group in a multi-arm pre-post study. These studies have the strength of temporality to be able to suggest that the outcome is impacted by the intervention, however, pre-post studies do not have control over other elements that are also changing at the same time as the intervention is implemented. Therefore, changes in disease occurrence during the study period cannot be fully attributed to the specific intervention. Outcomes measured for pre-post intervention studies may be binary health outcomes such as incidence or prevalence, or mean values of a continuous outcome such as systolic blood pressure may also be used. The analytic methods of pre-post studies depend on the outcome being measured. If there are multiple treatment arms, it is also likely that the difference from beginning to end within each treatment arm are analysed.

Non-randomized trial study design

Non-randomized trials are interventional study designs that compare a group where an intervention was performed with a group where there was no intervention. These are convenient study designs that are most often performed prospectively and can suggest possible relationships between the intervention and the outcome. However, these study designs are often subject to many types of bias and error and are not considered a strong study design.

Randomized controlled trial study design

Randomized controlled trials (RCTs) are the most common type of interventional study, and can have many modifications (26-28). These trials take a homogenous group of study participants and randomly divide them into two separate groups. If the randomization is successful then these two groups should be the same in all respects, both measured confounders and unmeasured factors. The intervention is then implemented in one group and not the other and comparisons of intervention efficacy between the two groups are analysed. Theoretically, the only difference between the two groups through the entire study is the intervention. An excellent example is the intervention of a new medication to treat a specific disease among a group of patients. This randomization process is arguably the largest strength of an RCT (26-28). Additional methodological elements are utilized among RCTs to further strengthen the causal implication of the intervention's impact. These include allocation concealment, blinding,

measuring compliance, controlling for co-interventions, measuring dropout, analysing results by intention to treat, and assessing each treatment arm at the same time point in the same manner.

Crossover randomized controlled trial study design

A crossover RCT is a type of interventional study design where study participants intentionally "crossover" to the other treatment arm. This should not be confused with the observational case-crossover design. A crossover RCT begins the same as a traditional RCT, however, after the end of the first treatment phase, each participant is re-allocated to the other treatment arm. There is often a wash-out period in between treatment periods. This design has many strengths, including demonstrating reversibility, compensating for unsuccessful randomization, and improving study efficiency by not using time to recruit subjects.

Allocation concealment theoretically guarantees that the implementation of the randomization is free from bias. This is done by ensuring that the randomization scheme is concealed from all individuals involved (26-30). A third party who is not involved in the treatment or assessment of the trial creates the randomization schema and study participants are randomized according to that schema. By concealing the schema, there is a minimization of potential deviation from that randomization, either consciously or otherwise by the participant, researcher, provider, or assessor. The traditional method of allocation concealment relies upon sequentially numbered opaque envelopes with the treatment allocation inside. These envelopes are generated before the study begins using the selected randomization scheme. Participants are then allocated to the specific intervention arm in the pre-determined order dictated by the schema. If allocation concealment is not utilized, there is the possibility of selective enrolment into an intervention arm, potentially with the outcome of biased results.

Blinding in an RCT is withholding the treatment arm from individuals involved in the study. This can be done through use of placebo pills, deactivated treatment modalities, or sham therapy. Sham therapy is a comparison procedure or treatment which is identical to the investigational intervention except it omits a key therapeutic element, thus rendering the treatment ineffective. An example is a sham cortisone injection, where saline solution of the same volume is injected instead of cortisone. This helps ensure that patients do not know if they are receiving the active or control treatment. The process of blinding is utilized to help ensure equal treatment of the different groups, therefore continuing to isolate the difference in outcome between groups to only the intervention being administered (28-31). Blinding within an RCT includes patient blinding, provider blinding, or assessor blinding. In some situations it is difficult or impossible to blind one or more of the parties involved, but an ideal study would have all parties blinded until the end of the study (26-28,31,32).

Compliance is the degree of how well study participants adhere to the prescribed intervention. Compliance or non-compliance to the intervention can have a significant impact on the results of the study (26-29). If there is a differentiation in the compliance between intervention arms, that differential can mask true differences, or erroneously conclude that there are differences between the groups when one does not exist. The measurement of compliance in studies addresses the potential for differences observed in intervention arms due to intervention adherence, and can allow for partial control of differences either through post hoc stratification or statistical adjustment.

Co-interventions, interventions that impact the outcome other than the primary intervention of the study, can also allow for erroneous conclusions in clinical trials (26-28). If there are differences between treatment arms in the amount or type of additional therapeutic elements then the study conclusions may be incorrect (29). For example, if a placebo treatment arm utilizes more over-the-counter medication than the experimental treatment arm, both treatment arms may have the same therapeutic improvement and show no effect of the experimental treatment. However, the placebo arm improvement is due to the over-the-counter medication and if that was prohibited,

there may be a therapeutic difference between the two treatment arms. The exclusion or tracking and statistical adjustment of co-interventions serves to strengthen an RCT by minimizing this potential effect.

Participants drop out of a study for multiple reasons, but if there are differential dropout rates between intervention arms or high overall dropout rates, there may be biased data or erroneous study conclusions (26-28). A commonly accepted dropout rate is 20% however, studies with dropout rates below 20% may have erroneous conclusions (29). Common methods for minimizing dropout include incentivizing study participation or short study duration, however, these may also lead to lack of generalizability or validity.

Intention-to-treat (ITT) analysis is a method of analysis that quantitatively addresses deviations from random allocation (26-28). This method analyses individuals based on their allocated intervention, regardless of whether or not that intervention was actually received due to protocol deviations, compliance concerns or subsequent withdrawal. By maintaining individuals in their allocated intervention for analyses, the benefits of randomization will be captured (18,26-29). If analysis of actual treatment is solely relied upon, then some of the theoretical benefits of randomization may be lost. This analysis method relies on complete data. There are different approaches regarding the handling of missing data and no consensus has been put forth in the literature. Common approaches are imputation or carrying forward the last observed data from individuals to address issues of missing data (18,19).

Assessment timing can play an important role in the impact of interventions, particularly if intervention effects are acute and short lived (26-29,33). The specific timing of assessments are unique to each intervention, however, studies that allow for meaningfully different timing of assessments are subject to erroneous results. For example, if assessments occur differentially after an injection of a particularly fast acting, short-lived medication the difference observed between intervention arms may be due to a higher proportion of participants in one intervention arm being assessed hours after the intervention instead of minutes. By tracking differences in assessment times, researchers can address the potential scope of this problem, and try to address it using statistical or other methods (26-28,33).

Randomized controlled trials are the principle method for improving treatment of disease, and there are some standardized methods for grading RCTs, and subsequently creating best practice guidelines (29,34-36). Much of the current practice of medicine lacks moderate or high quality RCTs to address what treatment methods have demonstrated efficacy and much of the best practice guidelines remains based on consensus from experts (28,37). The reliance on high quality methodology in all types of studies will allow for continued improvement in the assessment of causal factors for health outcomes and the treatment of diseases.

Standards of research and reporting

There are many published standards for the design, execution and reporting of biomedical research, which can be found in Table 3. The purpose and content of these standards and guidelines are to improve the quality of biomedical research which will result in providing sound conclusions to base medical decision making upon. There are published standards for categories of study designs such as observational studies (e.g. STROBE), interventional studies (e.g. CONSORT), diagnostic studies (e.g. STARD, QUADAS), systematic reviews and meta-analyses (e.g. **PRISMA**), as well as others. The aim of these standards and guideline are to systematize and elevate the quality of biomedical research design, execution, and reporting.

- **Consolidated Standards Of Reporting Trials** (CONSORT, www.consort-statement.org) are interventional study standards, a 25 item checklist and flowchart specifically designed for RCTs to standardize reporting of key elements including design, analysis and interpretation of the RCT.
- **Strengthening the Reporting of Observational studies in Epidemiology** (STROBE, www.strobe-statement.org) is a collection of guidelines specifically for standardization and